

## **English in Switzerland: Inherent variation in a non-native speech community**

*The study might show that we have our own special swissicized variety of english, at which point you can wonder if it isn't also a type of national language, which combines structures, expressions and mindsets of all the other national languages, une sorte de mélange. Whatever. (IFMSA speaker M, e-mail)*

***Thèse de Doctorat présentée devant la Faculté des lettres de l'université de Fribourg, en Suisse. Approuvé par la Faculté des Lettres sur proposition des professeurs Peter Trudgill (premier rapporteur) et Sali Tagliamonte (deuxième rapporteur). Fribourg, le 6 juin 2007. Le Doyen Jean-Michel Spieser***

Mercedes Durham, originaire de Leysin (VD).



**"Je déclare sur mon honneur que ma thèse est une oeuvre personnelle,  
composée sans concours extérieur non autorisé, et qu'elle n'a pas été  
présentée devant une autre Faculté"**

**Lieu et date:**

**Signature:**

**York, le 1 décembre 2006**



## Table of contents

Table of contents .....	2
Acknowledgments .....	5
<b>Part I – Introduction</b> .....	<b>7</b>
<b>1. Languages in Switzerland</b> .....	<b>7</b>
1.1 Introduction .....	7
1.2 The Linguistic Situation of Switzerland .....	9
1.3 The rise of English in Switzerland and its causes .....	14
1.4 Examples of English in Switzerland .....	18
1.5 Census Results .....	21
1.6 Consequences of the Swiss Linguistic Situation .....	28
<b>2. Background Literature and Aims</b> .....	<b>29</b>
2.1 Introduction .....	29
2.2 Language Contact .....	29
2.2.1 Focussing .....	30
2.2.2 Tertiary Hybridization .....	32
2.2.3 New Dialect Formation .....	34
2.3 English as a lingua franca .....	38
2.4 Aim 1: focussing in Switzerland .....	42
2.5 Comparing linguistic varieties .....	43
2.6 Language Variation and Change .....	44
2.6.1 Methodology of LVC .....	45
2.6.2 Multivariate analysis .....	47
2.7 Comparative Sociolinguistics .....	50
2.8 Second Language Variation .....	56
2.9 Second Language Acquisition .....	58
2.10 Aim 2: Non-native acquisition of variation .....	61
<b>3. Data Sources</b> .....	<b>63</b>
3.1 Introduction .....	63
3.1.1 Data Collection .....	63
3.1.2 Observer's paradox and non-native speakers .....	63
3.1.3 Aims of the present chapter .....	64
3.2 Swiss data .....	64
3.3 IFMSA-Switzerland .....	66
3.3.1 The Association .....	66
3.3.2 Languages Used in the Association .....	69
3.3.3 The Mailing List .....	71
3.3.4 The Situation at Meetings (and Interviews) .....	73
3.3.5 Data and Methods .....	75
3.3.6 Results .....	76
3.3.7 Change over Time .....	80
3.3.8 Causes of Change .....	88
3.3.9 Language Comprehension .....	91
3.3.10 Type of English used .....	94

3.3.11 Final words on IFMSA-CH .....	95
3.4 Native English data .....	97
3.5 Extra data.....	99
3.6 Discussion .....	100
<b>4. Methodology and variable selection.....</b>	<b>102</b>
4.1 Introduction .....	102
4.2 Feature selection and presentation .....	102
4.2.1 Feature selection .....	102
4.2.2 Presentation of features.....	107
4.3 Lexical Variation I: Informations.....	110
4.3.1 Introduction.....	111
4.3.2 Non-count nouns in English.....	112
4.3.3 Information in the source languages .....	113
4.3.4 Extraction and coding .....	115
4.3.5 Results and analysis .....	119
4.3.6 Discussion .....	126
<b>Part II.....</b>	<b>128</b>
<b>5. Future .....</b>	<b>128</b>
5.1 Introduction .....	128
5.2 Future Forms .....	128
5.3 Future forms in the three Swiss languages.....	132
5.4 Previous studies.....	134
5.5 Data sources .....	135
5.6 Extraction and Coding.....	135
5.7 Results .....	136
5.8 Discussion .....	144
5.9 Conclusion.....	146
<b>6. Relatives.....</b>	<b>147</b>
6.1 Introduction .....	147
6.2 Relative Clauses .....	148
6.3 Relative clauses in the source languages.....	152
6.4 Previous Studies .....	155
6.5 Data Sources.....	156
6.6 Extraction and coding.....	156
6.7 Results and Analysis .....	163
6.8 Multivariate analysis .....	177
6.9 Discussion .....	185
6.10 Conclusion.....	186
<b>7. Complementizers .....</b>	<b>187</b>
7.1 Introduction .....	187
7.2 Complementizers.....	188
7.3 Complementizers in the non-native languages.....	189
7.4 Previous studies.....	190
7.5 Extraction and Coding.....	192
7.6 Results .....	195
7.7 Multivariate analysis .....	211

7.8 Discussion .....	214
7.9 Conclusion.....	216
<b>8. Lexical Variation II: Also, as well and too .....</b>	<b>217</b>
8.1 Introduction .....	217
8.2 Also, as well and too .....	218
8.3 Additive adverbials in the source languages .....	220
8.4 Previous work on additive adverbials .....	221
8.5 Extraction and Coding.....	222
8.6 Results .....	223
8. 7 Discussion .....	230
8. 8 Conclusion.....	231
<b>Part III .....</b>	<b>232</b>
<b>9. Discussion and conclusion.....</b>	<b>232</b>
9.1 Introduction .....	232
9.2 Summary of overall findings.....	233
9.3 Implications of the results for aim 1: focussing in Switzerland .....	239
9.4 Implications of the results for aim 2: Second Language Variation.....	243
9.5 Additional Considerations.....	245
9.6 Closing words.....	246
References.....	248

## Acknowledgments

Most people would consider themselves lucky to have one inspiring teacher, so I must consider myself truly blessed to have had three of the very best. Thanks, first of all, to Peter Trudgill for showing me that sociolinguistics could be just as interesting as (if not more interesting than) literature and for much help and support with my thesis. I am so lucky to have had such a wonderful role model as my first linguistics teacher and as my supervisor. Secondly thanks to Sali Tagliamonte for her contagious enthusiasm for LVC and multivariate analysis and for always finding the time to help me when I needed it. The classes I took during my Erasmus year at York and the work I did as an R.A. gave me the skills I needed to apply LVC to the Swiss English project. Finally thanks to Jennifer Smith; not only was working with you on the kids data a blast, but it taught me the best way to present results. I appreciate the (subtle) nagging to make sure I got the thesis work done on top of my research.

Thanks also go to everyone who worked on the Fond National project; Professors David Allerton and Richard Watts, as well as Yvonne Droeschel, Babette Neukirchen and Lukas Rosenberger. Maybe our findings are still hazy in terms of focussing in Switzerland but we definitely found a lot of interesting things to say about Swiss English. Thanks also to staff and students at the English department in Fribourg and the Linguistics department in York, especially to my office mates, Regina, Didier and Denis in Switzerland and Hazel, Gareth, Leendert and Nanna in England. You were always ready for a chat if the need to procrastinate got too strong.

I gratefully acknowledge the Swiss National Science Foundation whose research grant made my thesis and the research project on English in Switzerland possible.

Especial thanks go to the people who made the data collection possible; first and foremost all the members of IFMSA-CH who kindly gave me their time (and their e-mails) and allowed me to 'use' their association for the project. Not only did you help other medical students by organizing exchanges for them, but you've also helped



linguists as well. Thanks for help in data collection also go to Emma Arnold and Matt Ainley for providing me with access to their club's e-mails to use as a native control group.

Thanks to Yvonne, as well as Gilles Brunner and Heike Pichler for help with the German/Swiss German translations. Thanks also to Lena Pasche for providing me with the National Census raw data so we could see for ourselves how Swiss people use English before the results came out officially.

Thanks to my all friends who must have listened to far more about English in Switzerland and second language variation than they had bargained for; special mentions go to Sam, Matt, Julia, Sarah and Mike who sat through numerous practice presentations along the way but yet always had a smile and good questions! Thanks to Heike for always being willing to commiserate with me when I needed someone to listen to my rants or problems!

Thanks to my dad and grandmother for both constantly asking me how my 'paper' was going. It'll be a relief to finally tell you it's done! Thanks to my mom for believing I could do it, but mainly because if you hadn't made sure I was trilingual by the age of 8 ('dragging me around the world performing linguistic experiments' as some have put it) I wouldn't be the linguist I am today, in layman's terms and or as an actual linguist.

Thanks to Andre (or Andrea or Dante) for support and criticism – and of course especially for the sibling rivalry that wasn't going to let you finish a SECOND doctorate before I got one!! If nothing else that made sure I didn't keep it dragging along too long! Finally thanks to Pete Dodd for being a calming influence on me when I got too crazy or stressed – it's not an easy job I'm sure.

## Part I – Introduction

### 1. *Languages in Switzerland*

#### 1.1 Introduction

This thesis presents a portion of the results of a research project (see Table 1.1) funded by the Swiss National Science Foundation. The project ‘Language Contact and Focussing: The Linguistics of English in Switzerland’ was set up to examine the phenomenon of English as a lingua franca in Switzerland. Taking as a starting point the fact that a number of studies had already found that English was used extensively in the country (Cheshire and Moser, 1994, Dürmüller, 2001, 2002, Watts and Murray, 2001), the main aim of the project focused on determining:

- a) whether the situation of English in Switzerland might have fostered a distinct variety of English,
- b) whether this variety was shared by speakers of English from the three main linguistic groups in the country (i.e. French, German and Italian).

If these were determined to be accurate and the English in Switzerland could be considered to be distinct from other varieties of English, then this new variety could be considered to be Swiss English.

<p><b>Language Contact and Focussing: The Linguistics of English in Switzerland</b></p> <p>A research project funded by the Swiss National Science Foundation. (July 2001 to June 2004)</p> <p>Project Coordinators:</p> <ul style="list-style-type: none"> <li>• Prof. Peter Trudgill, University of Fribourg</li> <li>• Prof. Richard J. Watts, University of Bern</li> <li>• Prof. David Allerton, University of Basel</li> </ul> <p>Research Assistants:</p> <ul style="list-style-type: none"> <li>• Yvonne Dröschel, University of Fribourg</li> <li>• Mercedes Durham, University of Fribourg</li> <li>• Babette Neukirchen, University of Bern (2001-2002)</li> <li>• Lukas Rosenberger, University of Bern. (2002-2004)</li> </ul>
---






Table 1.1: Swiss National Science Foundation Project

Previous studies considering dialects in contact (Trudgill, 2004, Kerswill, 2002) had revealed that in situations of intense contact several varieties could merge into a single new dialect. A basic assumption of the project was that the interaction of French, German and Italian speakers of English was as intense as in these previous studies and thus this interaction was likely to have affected the English spoken in Switzerland.

The English spoken by a group of French natives may contain aspects of interference from French, just as the English spoken by a group of German natives may show German influence. The English spoken by native French, German and Italian speakers together could, however, potentially contain interference from all three languages which would make it a distinct variety.<sup>1</sup> To prove that there is a distinctly Swiss variety of English and that this variety differs from other European non-native varieties of English, I will have to demonstrate that the three linguistic groups use the features considered in this thesis in the same way. This will demonstrate that it is a genuinely intranational form rather than a variety of English spoken by non-native speakers of only one language.

This thesis will, first of all, present the linguistic situation in Switzerland, which will allow me to discuss the extent of English use in the country. It will then introduce the methods used to establish correspondence between different yet possibly related language varieties. Finally, I will present analyses of several linguistic features so that the English used by non-native speakers in Switzerland can be examined.

The remainder of this chapter will present the linguistic situation of Switzerland, focussing particularly on English and will introduce a few of the reasons which have been put forward to explain why English has acquired so much importance in Switzerland. The lingua franca status of English in Switzerland will be considered to determine when English acquired this status and to what extent English is used in Switzerland.

---

<sup>1</sup> The process through which a variety of language acquires a distinct form is known as *focussing* (Le Page and Tabouret-Keller, 1985:201). It will be discussed in greater detail in chapter 2.

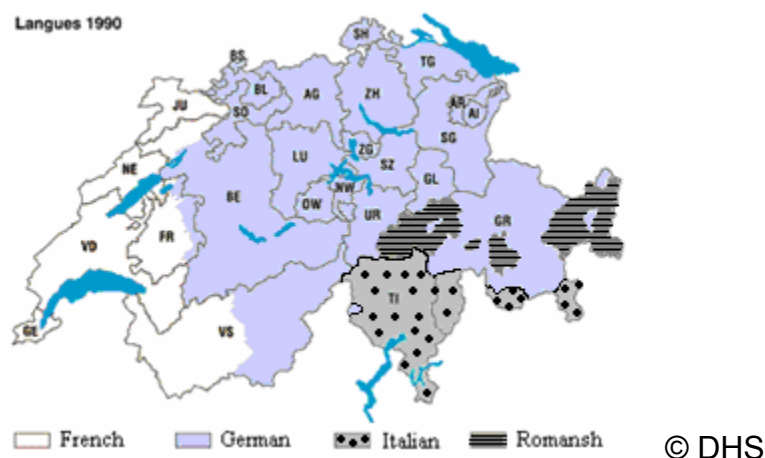
## 1.2 The Linguistic Situation of Switzerland

### a. Switzerland and its four languages

A country with slightly over seven million inhabitants, Switzerland has four national languages: German, French, Italian and Romansh and is therefore often mentioned in terms of its multilingualism and language policies (Edwards, 1995, Hoffman, 2000). Of these four languages, all but Romansh are considered to be official languages<sup>2</sup> and are used in the government and for federal administration. The official status of the three languages dictates that they should be given equal representation in the government and that any Swiss citizen should, in theory, be allowed to use any one of these languages for governmental or administrative purposes (Rash, 1998:30).

### b. Linguistic distribution

As a whole, Switzerland is multilingual but in actuality it is an amalgamation of distinct linguistic areas (see Figure 1.1).



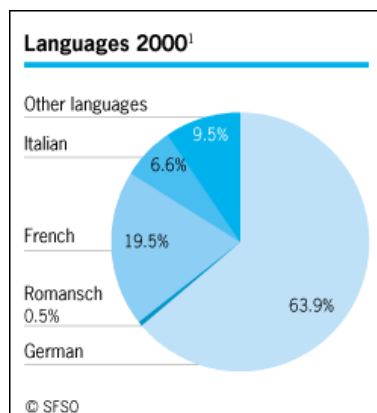
**Figure 1.1 Regional distribution of the national languages of Switzerland<sup>3</sup>**  
**[Map by Camillo Kohli. ©DHS, Dictionnaire Historique de la Suisse, 2003]**

<sup>2</sup> Romansh is the native language of only .5 % of the Swiss population (Swiss Federal Statistics Office (OFS), 2002). Although it has recently been given more status, it is not on a par with the other languages in terms of importance. Due to its extreme minority status, the research project and the present thesis did not examine Romansh speakers of English.

<sup>3</sup> The canton names are given in the table by two-letter abbreviations. From left to right: GE: Geneva, VD: Vaud, NE: Neuchatel, JU: Jura, BE: Berne, FR: Fribourg, VS: Valais, BS: Basel-Town, BL: Basel-Country, SO: Solothurn, AG: Aargau, LU: Lucerne, OW: Obwalden, NW: Nidwalden, SH: Schaffhausen, ZH: Zurich, ZG: Zug, SZ: Schwytz, UR: Uri, TI: Ticino, TG: Thurgau, SG: St Gall, GL: Glaris, AR: Appenzell-Outer-Rhodes, AI: Appenzell-Inner-Rhodes, GR: Grisons.

There is a French-speaking part of Switzerland, which is called *la Suisse Romande* by French speakers, a German-speaking part, called *die Deutschschweiz* by the German speakers and an Italian speaking part (restricted almost entirely to the canton of Ticino)<sup>4</sup>, called *la Svizzera italiana* by the Italian speakers. On the whole, the inhabitants of Switzerland live in monolingual communities (see Table 1.2 below for figures of this) but a number of cantons encompass two or more linguistic regions and there are bilingual towns and villages where most of the inhabitants use two languages daily. Fribourg in the canton of Fribourg (FR) and Bienne in the canton of Bern (BE) are two such bilingual towns.

Although the three national languages are considered to be equal, their relative distribution within the country is not. German has by far the greatest number of native speakers, with nearly two-thirds of the Swiss population (63.9 %) considering it to be their mother tongue at the time of the most recent national census in 2000 (see Figure 1.2 below).<sup>5</sup> French native speakers make up approximately 20% of the population of Switzerland, while Italian speakers account for just below 7% (Swiss Federal Statistics Office - hereafter OFS - 2002).



**Figure 1.2 Relative size of the national languages of Switzerland (reported mother tongue speakers) (OFS, 2002)**

<sup>4</sup> Italian is also spoken in some parts of the Grisons as can be seen on Figure 1.1. Romansh is spoken only in the canton of the Grisons.

<sup>5</sup> As will be discussed below, the census asked which language the respondent thought in and spoke the best. This can be interpreted as mother tongue, but may also apply merely to someone's main language.

Table 1.2 underlines the extent to which the linguistic divisions are maintained; in the German-speaking area, for example, less than 5 % of people speak a national language other than German as their main language.

Table 1.2: National and non-national languages by language area Census 2000 (OFS, 2002)							
language area	Resident population	Speaking the language of the linguistic area <sup>6</sup>		Speaking another national language		Speaking a non-national language	
	Total	In absolute figures	in %	In absolute figures	in %	In absolute figures	in %
Switzerland	7'288'010	6'208'918	85.2	422'553	5.8	656'539	9
German language area <sup>7</sup>	5'221'135	4'519'601	86.6	245'621	4.7	455'913	8.7
French language area	1'720'365	1'404'482	81.6	137'507	8	178'376	10.4
Italian language area	320'247	266'730	83.3	32'287	10.1	21'230	6.6
Romansh language area	26'263	18'105	68.9	7'138	27.2	1'020	3.9

As can be seen in Table 1.2,<sup>8</sup> except for the Romansh language area, the percentage of people speaking the language of the linguistic area is above 80%. Thus, broadly speaking, the main language of individual speakers is the same as that of the canton or region in which they live. The lower percentage found in the Romansh-speaking area (i.e. 70%)

<sup>6</sup> This table seems to have been mistranslated on the statistics website as the above line is presented as 'Speaking a language from another area' – the French version was the source of the translation above.

<sup>7</sup> The Swiss statistics website does not specify which of the bilingual cantons it deems to be German-speaking and which it deems to be French-speaking.

<sup>8</sup> It is important to note that most of the results in this chapter come either from the website of the Swiss statistics office or from figures provided to me by members of a cantonal statistics office and as such provide a far more reliable and complete picture of the Swiss situation than it would have been otherwise possible to obtain.

can be explained in two ways; firstly, through the fact that Romansh is used by such a small fragment of the overall population in Switzerland, and secondly, because the Grisons is the canton which displays the highest number of different linguistic regions. There are Italian, German and Romansh speaking areas in the Grisons.

Generally speaking, Switzerland operates with a concept of linguistic territoriality; regions that have a majority of French speakers have French as the local language and those which have a majority of German speakers have German and so on. This undoubtedly has contributed to the feeling that none of the languages is more important than the others.

Despite this dividedness, there will necessarily be times when the inhabitants of the various linguistic regions will need to communicate with each other. The figures presented above are based on the analysis of a question of the respondents' mother tongue, so we need to establish what languages are used in situations where there are native speakers of several different languages. Before focussing on that, however, we must consider the other languages spoken within Switzerland's borders as they will help provide a full picture of Switzerland's linguistic situation.

#### c. Other languages in Switzerland

Alongside the official Swiss languages, there are a number of other languages spoken in the country, even as mother tongues. A total of 9% of the respondents of the census considered none of the official Swiss languages to be their main language. The 2000 census reveals that, although German, French and Italian are the top three languages spoken in Switzerland, Serbo-Croatian, Albanian, Portuguese, Spanish, English and Turkish are all used more often as a main language than Romansh (Table 1.3, OFS, 2002).

Table 1.3: Top Ten Languages Spoken in Switzerland	
1 German	4'639'762
2 French	1'484'411
3 Italian	470'961
4 Serbian, Croatian	103'350
5 Albanian	94'937
6 Portuguese	89'527
7 Spanish	76'750
<b>8 English</b>	<b>73'422</b>
9 Turkish	44'523
10 Romansh	35'072

Although all these non-indigenous languages contribute to Switzerland's varied linguistic landscape, one is especially relevant to the linguistic situation of Switzerland as a whole: English. As demonstrated in Table 1.4, English is used as a main language by one percent of the population. This table uses the percentages found in Figure 1.2, but places them in order of significance and provides the percentage of people who do not consider one of the official Swiss languages to be their main one. It also separates English from the other non-indigenous languages to show the percentage of its use in Switzerland (OFS, 2002, 2005).

Table 1.4: Percentage of native speakers of each of the four official languages in Switzerland (based on Figure 1.2)	
Language	Percentage
German	63.9
French	19.5
Italian	6.6
Romansh	0.5
Other languages	9.5
(English)	1.0

The special place English has among the languages in Switzerland is not, however, due to its native speakers. Rather, its importance lies in its use as a non-native language. Indeed, many people use English regularly in Switzerland, either with native English speakers or with other non-native speakers.



English is one of the languages used by Swiss people to communicate with each other at a national level, as numerous advertisements found in English can attest to (Cheshire and Moser, 1994). It is a quintessential *lingua franca* in Switzerland; the Swiss use English as a common language among groups who do not speak each others' languages. This is quite different to how it is used by native speakers and helps explain why most Swiss people consider the variety of English they use to be fairly different from Standard English (Watts and Murray, 2001).

### 1.3 The rise of English in Switzerland and its causes

English is the language most readily used in circumstances when speakers of more than one of the Swiss languages have dealings with each other. The importance of English in Switzerland is relatively new, however; 'before World War II the presence of English could be felt only where the citizens of the Anglo-American countries—then above all Britain—came into contact with Swiss citizens' (Dürmüller, 2002:115). If English was used by Swiss nationals, it was only to communicate with English speakers and never with other Swiss people. This use of English with non-Swiss people still occurs to some extent today, as Switzerland places a large emphasis on tourism. However,

'since the end of World War II, particularly since about 1960, the situation regarding the place of English among the languages used in Switzerland has drastically changed ... People in Switzerland are now often more familiar with English than with the languages of their compatriots. English has definitely changed its status, it has moved from the fringes to the centre, from the status of a foreign language to that of an additional language with LINGUA FRANCA function, i.e. a language that can be used for special purposes and for wider communication' (Dürmüller, 2002:116, emphasis in original).

If we accept the date of 1960, then it is reasonable to assume that we are in the presence of the second generation of Swiss speakers of English. Note, however, that, quite

naturally, English competence is not passed from one generation to the next; parents do not speak in English with their children. The English in Switzerland is continuously being renewed and each generation is likely to have the same learning difficulties as their parents (i.e. the French speakers will make similar mistakes to their parents as they too will have interference from French). The youngest generations in Switzerland (those who would have grown up in the period where English is thought to have acquired more lingua franca status and those born from then onwards) will have been exposed to English much more than the older generation and are much more likely to use English.

There are three main reasons; neutral language choice, language learning and diglossia; which suggest why the shift to English occurred in Switzerland. These factors have influenced the spread of English as a lingua franca in Switzerland and might well be restricted to countries with a multilingual background such as Switzerland. Of course, the other more common factors influencing linguistic choice, such as the linguistic marketplace value of English (Bourdieu, 1991) for business and industry and the current world-wide importance of English (Crystal, 2003), play a role as well.

- Neutral language choice

The fact that the majority of Swiss citizens are monolingual is a problem given that the country itself is multilingual and there will be times people from other linguistic regions will need to communicate with each other. Although a sizeable portion of the population can speak two or more national languages, more than half of all Swiss people say that they are monolingual and are not competent enough to use another Swiss language. This is despite the fact that they have most likely studied another Swiss language at school (see Section 1.5 for the census results). Because of this, speakers from the various linguistic regions of Switzerland have had to find a different way of communicating with one another. English is on the school curriculum, just as the national languages are, and many former students have reported that they find English easier to use than the other languages which they acquired at school.

Moreover, a number of Swiss English speakers justified their use of English by claiming it was more fair, as choosing one Swiss language over another one (French or German)<sup>9</sup> would be giving an unfair advantage to the native speaker of the one selected. Using English puts everyone at an equal and democratic (dis)advantage and, as such, represents a more neutral language choice. As it is a foreign language for almost all Swiss people, the three groups are assumed to have an equal facility using it.

Furthermore, considering the sheer quantity of information provided in English in Switzerland for the benefit of foreigners, many Swiss people's contact with English is likely to be as significant as their amount of contact with other Swiss languages. If material is presented only in French and in English, a German speaker might look at the English version rather than the French. As well as being a neutral language choice, English can be considered a default language in Switzerland.

The idea of English as a default language can also be seen the general avoidance of having to state things twice. Presenting everything in both French and German demands much more effort on the part of the speaker, or more likely writer, than using only English. An example of this can be found in a medical student brochure that outlines the courses students need to take during their studies. The writers of the brochure decided to write it in English, rather than in French and German, and explain their decision by citing economy (i.e. it would reduce their workload).

‘The Working Group gladly accepted the suggestion of the Joint Commission of Swiss Medical Schools to prepare the Catalogue in English, **in order to reduce the workload**. English has replaced Latin as the international lingua franca of medicine; if needed, medical students in Switzerland are offered a course in medical English at the beginning of medical school’. (Swiss Catalogue of Learning Objectives for Undergraduate Medical Training, Bürgi et al., 2001:9, emphasis mine)

---

<sup>9</sup> Italian would not generally be chosen as the lingua franca in mixed language groups as only a relatively small proportion of people outside Ticino learn Italian at school.

- Language Learning

The second factor in the use of English in Switzerland is tied to language learning. Almost all Swiss school children learn some English nowadays. Some might only study it for a year or two, whilst others might have six but nearly all have some knowledge of English.

Until recently, the school system required that students' first foreign language be another Swiss national language, so that German speakers studied French (or occasionally Italian), French speakers German, and Italian speakers studied either French or German. In the past few years, however, English has become more important as a second language. In early 2004, the schools of the canton of Zurich (ZH) decided to make English the first foreign language taught. A number of other German cantons have followed suit. These decisions have received extensive press coverage in the Swiss media as they were seen by many French speakers to be a slight to their language; in both 2004 and 2005, *Le Temps*, one of the main French-speaking Swiss papers, devoted a large section to the decisions of the Swiss German cantons (Busslinger, 2004, Cossy, 2005)

Although the choice to teach English as the first foreign language in schools is a recent development, there is some evidence to suggest that English has always been used with greater competence than other second languages. Discussions during the interviews conducted as part of the research project, as well as anecdotal evidence, lead to the conclusion that many French speakers feel that they left school better equipped to use English than German, despite the fact that they may in fact have studied German for slightly longer. This was echoed in discussions with German speakers as well; English was easier than French for them. The reasons for this are difficult to guess, as this feeling could be due to concrete differences in language teaching, to the fact that French speakers find it easier to learn English than German, or simply that French speakers want to learn English but not German.

- Diglossia

The third factor which can help explain the spread of English in Switzerland has to do with the varieties of German which are spoken in Switzerland. Although German is one of the national languages, the majority of Swiss Germans speak a Swiss German dialect as their native language, not High German. The situation in the German-speaking part of Switzerland is one of diglossia, in that High German is used for writing and in formal situations, but a Swiss German dialect is used in informal situations (Ferguson, 1959, Rash, 1998:49ff). In the French-speaking part of Switzerland, schoolchildren are taught High German and not one of the dialects spoken by their compatriots. Because of this, despite having studied German, a national language, for up to eight years, Swiss French and Swiss Italian speakers are most likely not able to understand a conversation between two Swiss German speakers. Although Swiss German is closely related to Standard German, there are enough differences between the two to make it near impossible for a French or Italian speaker with only a basic knowledge of High German to understand it.

It is not surprising that English is found to be the language of choice for lingua franca use, given the increasing emphasis is put on its teaching and given that the Swiss French speakers who do learn German are learning a variety which is not their compatriots' native language.

#### **1.4 Examples of English in Switzerland**

A few examples of the way English is used in Switzerland will be presented here, although the main presentation of potentially Swiss English features will be dealt with in the analysis chapters. This section will provide an idea of how prevalent the use of English is in the country as well as giving a few intimations of the type of English used.

Some of the advertisements in English found in Switzerland merely demonstrate Standard English use. But others, like those below, serve to show how English can be used playfully (by translating French idioms word by word) in the case of the Swiss post

office advertisements (figures 1.4 and 1.5), or how it is used in a clearly non-standard way, as in the case of the banking advertisement (figure 1.3).

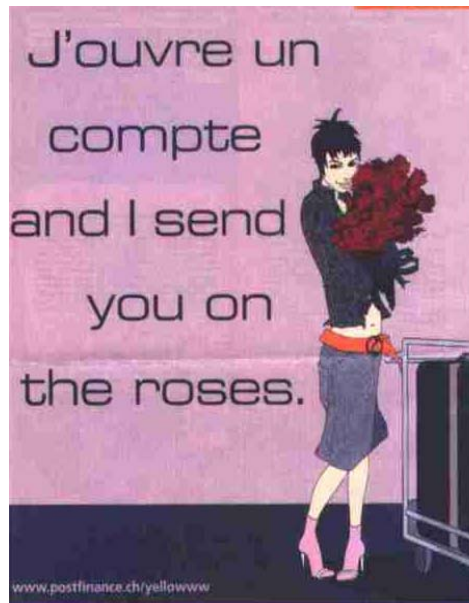
Figure 1.3 Advertisement which appeared several times in the *International Herald Tribune* in 2002

ARE YOU INTERESTED TO REACH THE SWISS FINANCIAL  
COMMUNITY?

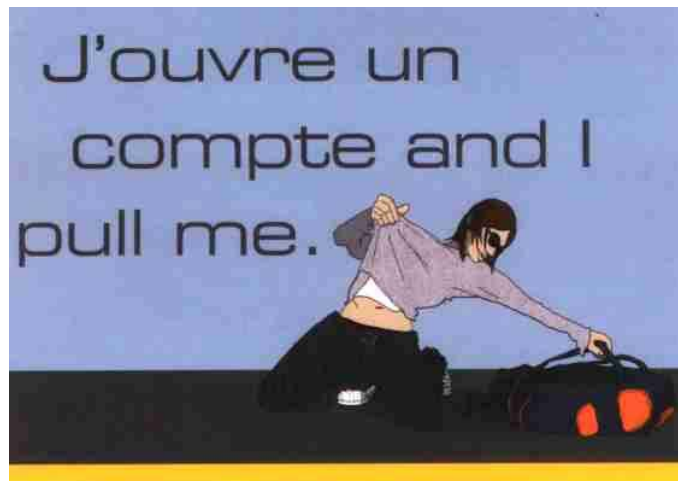
We represent, in Switzerland, since 10 years X's Fund Services (ex X)

Consequently, we are in partnership with most of the Swiss Bankers and Mutual  
Fund Managers.

If you have a service or a product which ought to appeal to the Swiss Financial  
Industry, we have the know-how, the skills and the right introductions to  
represent you in Switzerland.



Figures 1.4-1.5 (Swiss Post advertisements, with a play on French idioms, 'envoyer sur les roses', 'je me tire')



## **1.5 Census Results**

As the primary aim of the research project was to establish whether there was a single Swiss English variety, the whole of the country has to be considered at least briefly to determine whether any specific findings in one Swiss group could be extended to the overall situation. The findings of the most recent national census allow us to establish how much English is used in Switzerland by the general population.

The results of this part of the study come from an analysis of the 2000 Swiss National census. As with the distribution of main languages presented in the first section, the census also reveals a number of facts about what languages Swiss speakers use in addition to their native language. The census results also underline the degree to which Swiss citizens are for the most part monolingual and highlight a fact that many journalists, academics and laymen have sensed for years: the growing importance of English as a language in Switzerland, for international, commercial reasons, but also as a tool for intranational communication (Dürmüller, 2002:116, Watts and Murray, 2001).

Although the full census results have not as yet been released by the Swiss Statistics Office, the results for the main languages, (see Figure 1.2 and Tables 1.1 and 1.2) have been compiled and released to the press. The unreleased results of the census presented below (Table 1.6 and figures 1.6 and 1.7) were compiled by myself, based on the raw figures from the census provided by the statistics office of the Canton of Vaud.

Questions 8a and 8b of the 2000 census deal with language use (see Table 1.5 and 1.5'). Although the first question allows only a single answer, the second question allows for multiple responses.



[illegible]

Table 1.5: Questions on language use as presented in the 2000 Swiss census (OFS, 2002)

[illegible]

Table 1.5' – Translation of Table 1.5

By adding up the number of people who reported that they used English in the various categories in the data provided by the statistics office, I obtained the percentages of English use in Switzerland for the country as a whole and for each individual canton. The results are presented in Table 1.6 and Figure 1.6; the table gives the figures grouped by the linguistic divisions of the country (French, German, Italian + Romansh), whereas the figure plots English use in Switzerland from the canton with the highest percentage of

users to the lowest. The percentage of self-assessed completely monolingual speakers (of any language) for Switzerland and each canton was also calculated to determine whether a high percentage of English use could be correlated to cantons more open to other languages. The results for this are presented in Table 1.6 and in Figure 1.7.

The cantons have been grouped broadly by linguistic background on Table 1.6 in order to establish if there are any differences by linguistic group. Of the bilingual cantons, Berne is placed in the German-speaking category and Fribourg and Valais in the French-speaking area. The cantons of Grisons (Romansh, German and Italian speaking) and Ticino (Italian speaking) have been placed between the German and the French cantons.<sup>10</sup>

The figures given in Table 1.6 also present the overall percentage of foreigners in each canton to determine whether this might have affected the overall distribution of English. These results were available from the OFS website ([www.statistik.admin.ch](http://www.statistik.admin.ch)). This was done on the assumption that a canton with a high number of immigrants was likely to be more multilingual than one which does not. To a certain extent, the percentage of non-Swiss citizens in an area also provides an indication of English use. A large portion of the immigrants in the cantons of Geneva, Zurich and Basle are researchers and employees in international companies and foundations and as such are likely to use English.

Table 1.6: Percentage of English use in Switzerland by canton			
Canton	Percentage of reported English use	Percentage of monolingual speakers	Percentage of foreigners (from OFS website)
Zurich	21.8	51.5	22.1
Berne	13.6	60.6	12.2
Lucerne	11.3	66.7	15.4
Uri	6.5	78.3	8.0
Schwytz	12.3	67.9	15.6
Obwalden	9.6	73.1	11.0
Nidwalden	13.9	69.9	9.5
Glaris	7.8	66.7	20.1
Zug	21.5	54.9	19.7

<sup>10</sup> The division by linguistic background was a decision on my part rather than on the part of the Statistics Office which had not used a particular grouping for the cantons.

Solothurn	11.2	63.9	17.1
Basle-Town	19.6	47.5	27.8
Basle-Country	17.1	58.8	17.5
Schaffhausen	12.7	62.8	20.0
Appenzell-Outer-Rhodes	10.4	69.7	13.7
Appenzell-Inner-Rhodes	8.0	79.9	10.2
St Gall	10.7	65.9	19.9
Aargau	14.1	62.0	19.3
Thurgau	10.4	66.2	19.1
Grisons	9.9	55.8	13.2
Ticino	6.7	65.9	25.5
Vaud	13.2	56.0	26.9
Valais	6.9	65.4	16.6
Neuchâtel	7.9	60.8	22.8
Geneva	18.3	44.2	37.6
Jura	3.8	71.8	12.2
Fribourg	8.4	60.3	14.4
Switzerland	14	59.1	

Figure 1.6: Percentage of reported English use in Switzerland (by canton)

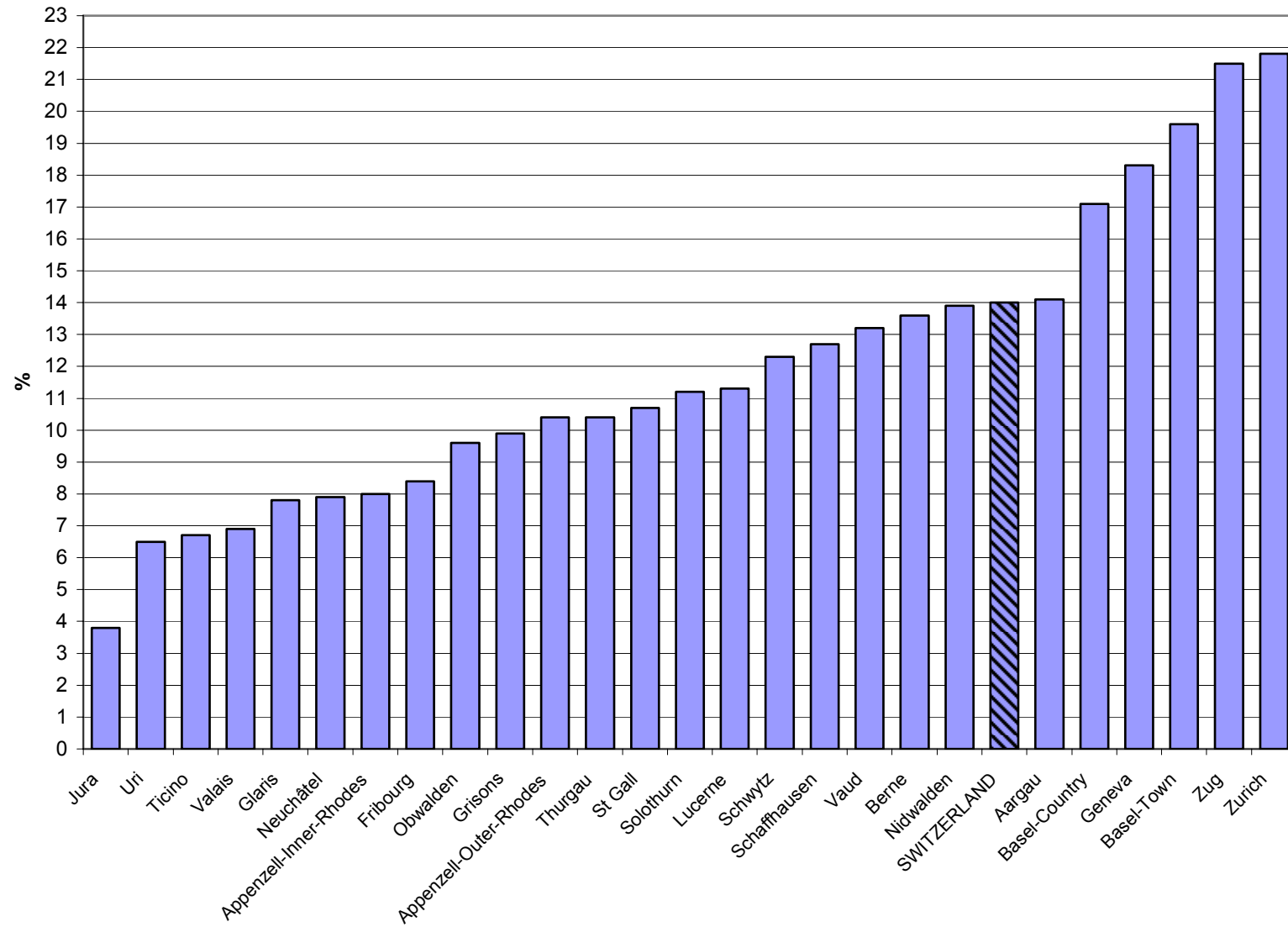
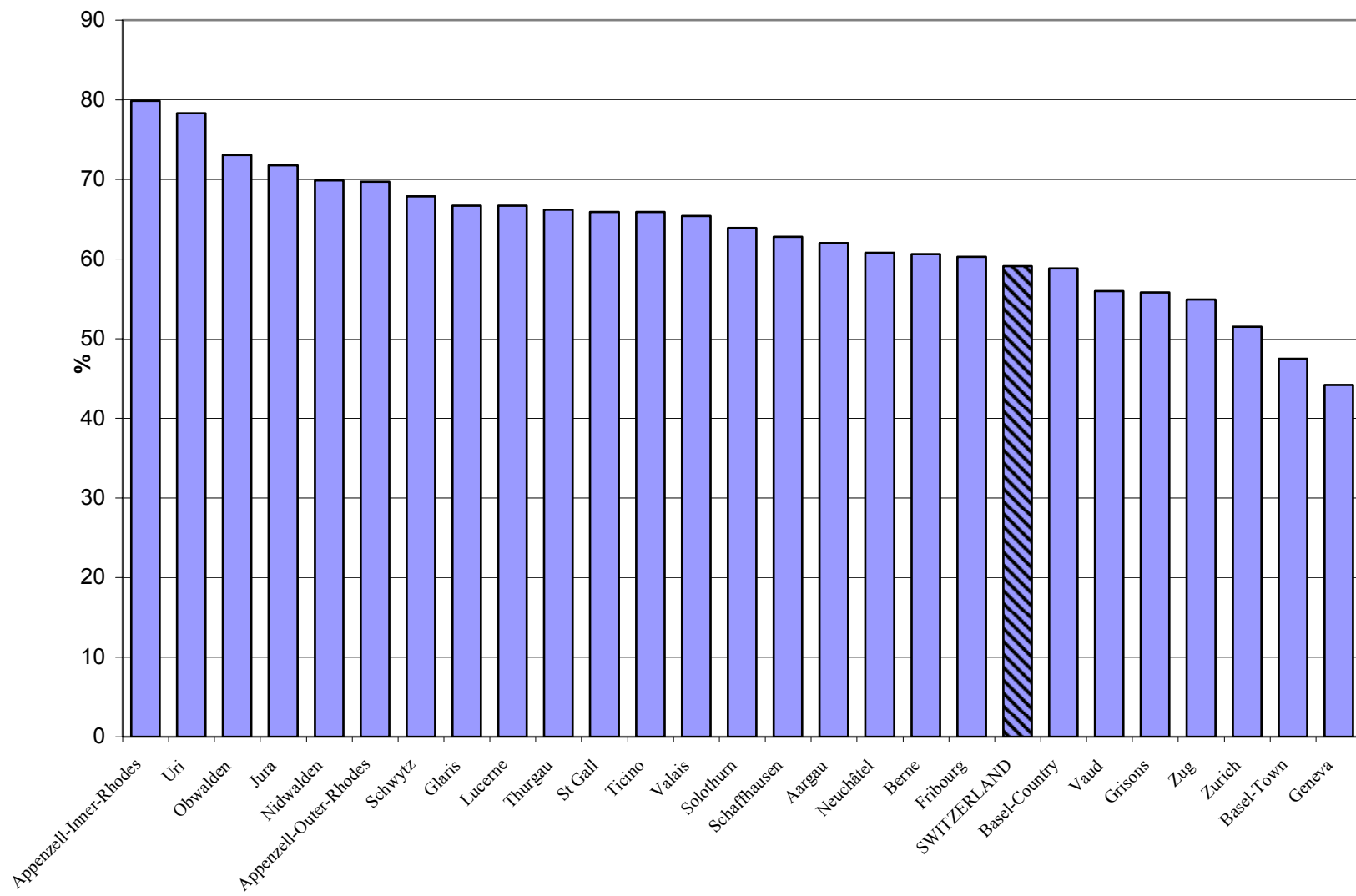


Figure 1.7: Percentage of monolingualism in Switzerland (by canton)



## Discussion of the results

Overall, fourteen percent of the Swiss population reported that they used English either at home or at work. This is a sizeable proportion of the population, but note that the rate varies from canton to canton. Compare Jura, the canton with the lowest figures of English use, at 3.8 percent, to Zurich, the canton with the highest figures at 21.8 percent. The differences between cantons can be explained by how much a canton can be considered rural or urban. For example, it is not surprising that Geneva is the French-speaking canton with the highest level of English users, as the city of Geneva has a large number of banks and foreign companies. Moreover, Geneva is one of the main seats of the U.N., and accordingly its citizens are from a large number of countries. The same explanation applies to Zurich and to some extent Zug, which is a tax haven and as such the base for a number of international companies.<sup>11</sup>

Although 14% English use is considerable, especially given that it is not an indigenous Swiss language, the results of the census most likely contain some degree of underreporting as well. First of all, the question on languages asked which languages were *generally* used, leaving open a certain degree of interpretation, as to what *generally* might mean.

Secondly, the census requested that students not note the languages they were in the process of learning. The results do not therefore include the language use of any of the population aged 12 to 18 who would fall into the category of English-language learners rather than users. Yet they are precisely the segment of the population that is the most likely to use English (even outside of school).

The high level of monolingualism uncovered in the census results (figure 1.7) is undoubtedly important to consider for the subsequent analysis.<sup>12</sup> Even in cantons that are divided linguistically (Fribourg, Valais and Berne for example) about 60% of people only use one language regularly. Another finding is that a number of Swiss

<sup>11</sup> Note that the census does not distinguish between people who have Swiss citizenship and people who merely live in Switzerland.

<sup>12</sup> The term *monolingual* as used in the census reports encompasses speakers who are in fact in a situation of diglossia (Ferguson, 1959), as for the purposes of these results Standard German and Swiss German are considered to be one entity. The same holds for the dyads of Standard Italian/Ticinese dialects and the less frequent French/Patois.

German speakers do not report any High German use but instead only use Swiss German (see Rash, 1998 for a further discussion of this). This is important in view of the fact that French speakers learn High German at school and not Swiss German.

This monolingualism, like the distribution of English shown in figure 1.6, varies according to the canton considered; the cantons which are on the border of a linguistic division (between the French and German linguistic regions primarily) demonstrate a higher degree of multilingualism. The cantons which the census revealed to have the highest percentages of monolingual speakers and low percentages of English use compared to the country as a whole tend to be the ones which are surrounded by cantons which share their linguistic background (Schwytz (SZ) and Obwalden (OW) on figure 1.1, for example).

## **1.6 Consequences of the Swiss Linguistic Situation**

This preliminary research into the linguistic situation in Switzerland supports the more central research aims of the thesis. The census results reflect the growing importance of English as a lingua franca in Switzerland and helped demonstrate that English is used by a significant proportion of the population for inter-regional contact. The census also established the extent to which Switzerland, though multilingual as a whole, is a country made up of many monolingual speakers.

The remainder of this thesis will focus on the internal consequences of this linguistic situation. It will attempt to establish what the English used in Switzerland is like and, more importantly, whether the different linguistic groups can be shown to be using a single variety.

## **2. Background Literature and Aims**

### **2.1 Introduction**

Recall from the previous chapter that one of the primary aims of the Swiss National Science Foundation research project was to uncover whether focussing was underway in the English used in Switzerland. Focussing can be defined as ‘the process by means of which [a] new variety acquires norms and stability’ (Trudgill, 2004:88). The project hypothesized that the linguistic situation in Switzerland could lead to the use of a single variety of Swiss English spoken by people from the various linguistic backgrounds. If the English in Switzerland can be shown to be a single variety, then focussing has taken place. In this section I discuss in more detail the terms ‘focussing’, ‘tertiary hybridization’ and ‘new dialect formation’ and their application to the current study. This chapter will also offer a novel approach to language contact phenomena. Specifically, it will demonstrate how the analysis of inherent variability in non-native speakers of English can be used to determine shared patterns across linguistic groups.

The situation of English in Switzerland is echoed in other situations of language contact and in the growing importance of English as a world language. Section 2.2 introduces a number of important aspects of language contact, particularly the notions of focussing (Le Page and Tabouret-Keller, 1985), tertiary hybridization (Whinnom, 1971) and new dialect formation (Kerswill, 2002, Trudgill, 2004), while section 2.3 deals with World English (Crystal, 2003, Kachru, 1982). The later sections of this chapter deal with the aspects of Language Variation and Change (Chambers, Trudgill and Schilling-Estes, 2002) and Second Language Acquisition (Ellis, 1985, 1994) that played a part in the elaboration of this thesis.

### **2.2 Language Contact**

The National Science Foundation Research project aimed to uncover whether a unified variety of English showing syntactic and lexical influence of the input languages existed in Switzerland. The English language contact between the three Swiss linguistic groups would have been a trigger for the creation of this variety. To



demonstrate how this is a possibility, this section will introduce three aspects of language contact which have been found to lead to the formation of new language varieties or new dialects. If these aspects are at play in Switzerland as well, then there is a distinct possibility that the same processes might have taken place.

### **2.2.1 Focussing**

Le Page and Tabouret-Keller (1985) developed the notion of focussing to account for processes found in the formation of new language varieties. In their view, the process through which a language form acquires a series of fixed norms through intensive contact is considered to be focussing. For Le Page and Tabouret-Keller (1985:201), an unfixed language variety, an ‘instant pidgin’, which is ‘diffuse, opportunistic, [and] involves all kinds of contextual cues to convey meaning’ can in the right circumstances become ‘more highly focussed [and] highly regularized’. In this focussed variety, ‘the functions of words and the relational functions between words have been grammaticized’ (Le Page and Tabouret-Keller, 1985:201).

Le Page and Tabouret-Keller (1985) applied the notion of focussing primarily to the development of creoles from pidgins. Pidgins are languages which arise through the contact of speakers of two or more languages using a language that is native to none of the groups. A pidgin is considered to be a language with no native speakers, and a creole is a pidgin which has acquired native speakers (Todd, 1991:48ff).<sup>13</sup> In the elaboration of a pidgin into a creole, the pidgin goes through a considerable amount of change. While the syntax of pidgins is seen as relatively volatile and unfixed, creoles are thought to have a more structured syntax. It is through focussing that a pidgin restricts its irregularity into a structured form.

One of the examples Le Page and Tabouret-Keller (1984:158) give comes from their analysis of the story-telling of four children from the Cayo District in Belize. All four children demonstrate different degrees of Belizean Creole and Standard English in their speech. However, rather than their speech being on a ‘linear scale from less Standard to more Standard, or from more Creole to less Creole, each child produced

---

<sup>13</sup> This is a somewhat simplified explanation of the difference between pidgins and creoles.

its own unique set of datum-points in relation to [...] external models' (Le Page and Tabouret-Keller, 1984:160). Although Le Page and Tabouret-Keller (1984:202-203) deem that the 'four children do not represent a highly focussed language community' and that 'any description of their community language will therefore have to be polysystemic', they note that 'the shared norms of the Cayo District are, under social pressures, evolving towards a more focussed shared system.'

Note that, although Le Page and Tabouret-Keller (1984:203) discussed focussing as a process, they did not study the specific ways in which a diffuse variety becomes a focussed one, stating that 'what happens in contact situations may be 'explained' *post hoc* (but not predicted)'. They recognize several 'agencies which promote focussing, however. These are:

- (a) close daily interaction in the community
- (b) an external threat or any other danger which leads to a sense of common cause
- (c) a powerful model – a leader, a poet, a prestige group, a set of religious scriptures
- (d) the mechanisms of an education system' (1984:187).

The notion that 'close daily interaction' is influential to focussing is tied to research on social networks (see Milroy and Milroy, 1985 and Milroy, 2002). For Le Page and Tabouret-Keller, it is only in situations where there was a considerable amount of contact between speakers and a general identification with a group that focused forms of language were likely to arise. Whilst these agencies are not all applicable to the Swiss situation, the notion of 'daily interaction' is. One of the findings of the census was that many Swiss people used English on a frequent basis, especially at work. We will return to the four agencies in chapter 3 when considering the specific group studied in this thesis, as there are a number of other similarities.

Alongside Le Page and Tabouret-Keller's work on focussing involving pidgin and creole varieties of language, more recently the notion of focussing has also been extrapolated to situations of new dialect formation (Kerswill, 2002:679, Trudgill, 2004:88). New Zealand English, for example, is distinct from other native varieties of English despite the fact that it has as its origins a number of British Isles dialects

(Trudgill, 2004:13). The aspects of focussing associated with new dialect formation will be dealt with in section 2.2.3.

As noted, much research has been conducted into creoles and stable non-native varieties of English once they have acquired a focused form, as in the case of the English spoken in India (Kachru, 1982). There is also a considerable amount of research into focussing in terms of new dialect formation (both completed and in progress) (see Trudgill, 2004 and Kerswill and Williams, 2000). To date, however, there has been no quantitative research into the processes involved in the focussing of a non-native language variety.

Focussing is an important notion for the Swiss situation for two reasons. First, as discussed in chapter 1, English has only been used as a lingua franca in Switzerland for the past forty years and we know from the census that it is only very recently that this use has become widespread. Moreover, English is used predominantly by speakers of the three main languages of the country (French, German and Italian) with one another, rather than with native English tourists (as discussed in the previous chapter). This widespread English use could have triggered the focussing of the three national forms (English as spoken by French speakers, English as spoken by German speakers and English as spoken by Italian speakers) into a single variety of Swiss English.

Secondly, if the English in Switzerland is found to be in the process of focussing, then it will be in the very incipient stages of its structuring. This may allow us to study the processes of focussing as they take place.

### **2.2.2 Tertiary Hybridization**

A second notion that can help understand the linguistic situation in Switzerland is the process of tertiary hybridization. Whinnom (1971) took the notion of tertiary hybridization from the field of biology and reinterpreted it in terms of the creation of new language varieties. The notion of linguistic tertiary hybridization attempts

primarily to account for the linguistic restructuring (i.e. how diffuse forms become a focussed and structured grammar) found in creoles.

For Whinnom (1971:105-6), a new pidgin is ‘unlikely to arise in a stable form’ in the communication of two speakers (or two groups of speakers) when it is the native language of one of the speakers, i.e. ‘a French-‘based’ pidgin [... in] the communication of an English-speaking schoolboy with a French schoolboy.’ Rather, it is in cases where the two speakers (or groups of speakers) are using a third language which is native to neither that a new form may arise, i.e. ‘a French-based pidgin could easily arise from the communication of an English schoolboy with a German schoolboy’ (Whinnom, 1971:106). The essential difference is that ‘in the second [situation], the target-language, French, is removed from consideration’ (Whinnom, 1971:106). In Whinnom’s (1971:106) view new language varieties can arise only in cases where there are two or more language groups speaking a third language, because in these cases:

‘Neither speaker has any model on which to improve his performance in French, nor any motive to improve it. Their motive can only be improved communication with each other, the inhibitions of the native who fears ridicule from speakers of the target-language are removed, and one can easily see how and why certain essential pidgin features could be produced.’

The relevance of this for the putative development of a single variety of Swiss English is clear. The English spoken in Switzerland is, in many cases, taught and spoken by non-native speakers to other non-native speakers. There is some interaction with native speakers but this is the exception rather than the norm for most Swiss people. This might mean that certain structures and forms might arise which would not develop in simple, monolingual second language acquisition. For example, Whinnom (1971:106) speaking of his German-English schoolboy French pidgin states that;

‘though both would have acquired and would maintain phonemes foreign to their native systems, their mispronunciations of French words, intelligible to each other and reinforced by repetition, might soon render the ‘French’ unintelligible to a native speaker’.

Although we would not expect such extreme cases such as Whinnom's example in the Swiss situation, there is a distinct possibility that lexical items, pronunciations and syntactic forms might undergo tertiary hybridization.

The concept of tertiary hybridization is relevant to the Swiss situation in cases where we find features which do not belong to native English but which exist in some form in the three source languages. Features found in the English of all three linguistic groups but stemming from only one of the source languages could also signal tertiary hybridization. So, for example, the French speakers might be found to use non-native forms transferred from German or Italian interlanguage features.

### **2.2.3 New Dialect Formation**

If one considers that the Swiss speakers used (or still use) an English influenced by their native languages, then we need to examine processes of new dialect formation to ponder how this too might help us uncover focussing in Switzerland. Moreover, as discussed above, some of the processes involved in new dialect formation have been shown to be similar to those of creolization and the formation of new non-native varieties.

Giles (1973) found that when speakers of different dialects interact with each other, they may accommodate their speech to that of others. That is to say they may make it more similar to the speech of the person they are speaking to. When this accommodation is especially intense (generally over an extended period of time) the speakers' own variety may change permanently, as long-term contact can alter speakers' variable rules (Coupland, 1984:65). In cases of transported varieties (either to a new country, as in the case of Australia and New Zealand (Trudgill, 2004) or more simply to a 'New town', i.e. Milton Keynes (Kerswill, 2002), the interaction between speakers of different dialects is especially intense. In cases such as these, the younger generations, i.e. those born within the new communities, find themselves in a situation where there is no stable community-wide variety for them to acquire so often they are the instigators of one.

The studies into new dialect formation can provide insight into the Swiss English situation on two counts. First of all, they focus on how features from the source dialects are selected to become part of the new dialect and secondly they discuss the time period and stages required for a new dialect to coalesce. New dialect formation has a number of processes in common with creole formation. It is in its examination of the stages of new dialect formation rather than the focussed new variety that this research can offer much to our understanding of how the English in Switzerland might be changing. The stages and processes found in new dialect formation will be presented below.

Linguistic contact in situations of new dialect formation is very intense. This intense contact allows several distinct dialects to merge into a single focussed variety. Note that a merger takes place rather than one dialect replacing the other ones. Variants found in different dialects can be levelled, leaving the new variety with a completely different distribution of variants than in any of the initial dialects. This was the case of New Zealand English where some of the first native-born speakers had features which were from several British dialect sources (Trudgill, 2004:104). One of the speakers Trudgill (2004:107) examined had ‘a phonological system’ which combined ‘features of Irish and/or Scottish origin with features that [were] obviously of English English origin,’ so that for example ‘his H dropping is combined with the presence of the distinction between /ʌ / and /w/’. Levelling is as a process through which disparate features are reduced but also a process through which forms distinctive to the new dialect are created (Kerswill, 2002:672).

As well as levelling, which he defines as ‘the loss of demographically minority variants’, Trudgill (2004:84-89) outlines five further processes involved in new dialect formation; mixing, unmarking, interdialect development, reallocation and focussing. Mixing is ‘the coming together in a particular location of speakers of different dialects of the same language, or of readily mutually intelligible languages’ (Trudgill, 2004:84). Unmarking ‘can be regarded as a subtype of levelling’ and Trudgill (2004:85) notes that ‘unmarked and more regular forms may survive even if they are not majority forms.’ Interdialect development has to do with ‘forms which were not actually present in any of the dialect contributing to the mixture, but which

arise out of interaction with them' (Trudgill, 2004:86). These interdialect forms are of three types; 'forms which are simpler or more regular than any of the forms present in the dialect mixture', intermediate forms or 'forms which are the result of hyperadaptation' (Trudgill, 2004:86-87). Trudgill's (2004:88) definition of focussing follows Le Page and Tabouret-Keller (1984); it is 'the process by means of which the new variety acquires norms and stability.' He notes that it 'is not to be identified with levelling. Although focussing implies levelling, the reverse is not the case: a reduction in the number of variants does not in itself lead to stability and societally shared norms' (Trudgill, 2004:88-89).

Trudgill (2004:89) adds that the processes of mixing, levelling, unmarking, interdialect development and reallocation 'can collectively be referred to as *koinéisation*. Koinéisation plus focussing constitute *new dialect formation*.' These processes do not all occur at the same time, but rather at different stages in the elaboration of the new dialect. Both the research in Milton Keynes and in New Zealand has revealed that new dialect formation takes place in several stages and through a number of generations.

It is the children of the first incomers to the new dialect area who are crucial for the formation of the new dialect as they do not have a predetermined model to acquire. Unlike children who acquire their native dialect in an environment where the dialect is relatively fixed, all the children in a new town, such as Milton Keynes, have to restructure the mix of dialects available into a levelled form. The dialect variety is still relatively unstructured at this stage as different speakers select different variants in their speech. A distinctive and focused dialect, considered to be the third stage of new dialect formation, only comes into being in the third generation of speakers (i.e. the second generation of 'native' born children) who, using further processes of focussing and levelling, produce a uniform language variety. This new variety has variability, of course, but often the patterning of the variants is different than it was in the source dialects. Note that the research of both Trudgill (2004) and Kerswill (2002) has focussed primarily on phonological data, so although they concern themselves with new *dialect* formation, it is in particular examples of *accent* that they give in their research.

Table 2.1 below, drawn from the discussion in Trudgill (2004:82ff) and Kerswill (2002:679), schematizes the different stages and main people involved in new dialect formation. Only the second generation of children born in New Zealand are considered to be speakers of a distinctly new dialect, New Zealand English, although the older generations of speakers show some degree of levelling.

Stage	Processes	Speakers involved
Stage I	rudimentary levelling and interdialect development	adult speakers of different varieties (Trudgill, 2004:83-99)
Stage II	extreme variability and apparent levelling in new-dialect formation	children (Trudgill, 2004:100-112)
Stage III	focussing, levelling and reallocation - determinism in new-dialect formation	second generation children (Trudgill, 2004:113-128)

Table 2.1 Stages of new dialect formation

In terms of the Swiss situation, the notion of new dialect formation is quite useful as Switzerland is a site of contact of different varieties of English. The notions discussed above will come into play if the English in Switzerland has acquired (or is acquiring) a focused form. The fact Switzerland is also involved in a considerable amount of direct language contact (French, German, Italian and English) is likely to have influenced some aspects of English variety contact.

These notions will however need to be interpreted with caution in terms of the Swiss situation; whereas the situation of New Zealand and of the ‘new towns’ in England rely quite heavily on direct transmission from one generation to the next (as presented in table 2.1) the Swiss situation does not have such language transmission. In Switzerland, English is transmitted through schooling and not from parent to child or through ‘peer acquisition’. Despite the fact that English has been a lingua franca in Switzerland for nearly 50 years (following Dürmüller’s calculations in 2002), it is quite probably still only in the incipient stages of focussing given that each generation starts anew in the learning of English. Many of the processes discussed above are thus unlikely to be fully completed.

In addition to processes which account for the formation of new linguistic varieties, we have to consider some of the reasons for the underlying importance of English in Switzerland and more generally in Europe and the world.



### 2.3 English as a lingua franca

As early the eighteenth century, English was mentioned as being in the process becoming a world-wide language:

“English will be the most respectable language in the world and the most universally read and spoken in the next century, if not before the close of this one.”

(President John Adams, 23 September 1780, from Kachru, 1982:2)

English has spread from being a language spoken within the confines of (part of) the British Isles to a language natively used across the world by millions of speakers (Crystal, 2003). The specific processes involved in this spread have been examined in detail; there are numerous studies which aim to present and describe the wealth of English varieties to be found world-wide (Gorlach, 1998 and 2002, Cheshire, 1991, Trudgill and Hannah, 1982, McArthur, 1998, etc). The spread of English as a native language was largely due to colonization. From the 16<sup>th</sup> century onwards, the British settled many continents, including North America and Australia, mostly for trade purposes. In these cases, the indigenous languages were lost in favour of English and current estimates put English as a native language of 400 million speakers (Crystal, 2003:67).

It is by no means only the native speakers of English who have ensured its place as one of the most widely spoken languages in the world (McArthur, 1998:355), however. Non-native speakers have made English a truly international language in the 20<sup>th</sup> and 21<sup>st</sup> centuries. Not all of the countries affected by British colonization have English as their first language, but instead it was established as an important second or official language in some. This is the case of countries such as India, Cameroon, Nigeria and Jamaica. These second language speakers of English are thought to make up another 430 million speakers (Crystal, 2003:65).

On top of these first and second language speakers of English, it is estimated that 750 million people speak English as a foreign language with at least a ‘medium level of

conversational competence' (Crystal, 2003:68). In all, 25% of the world's population speaks English with some level of competence (Crystal, 2003:69).

Alongside the studies which deal with native varieties of English, there are also numerous studies which attempt to account for the 'other side of English', the side which 'is concerned with English as the 'other tongue', or as a second language' (Kachru, 1982:1). These works consider that the purposes to which English is put have an important bearing on the type of English used and whether English can be considered a native language, a second language or a foreign language within a country.

To account for the differing forms of English in the world, Kachru (1982:38ff) developed a system of categorization, depending on the functions of Englishes and the 'models' which they follow. In his categorization, there are three different circles of English-speaking countries; inner, outer and expanding circles.

- 1) The **inner circle** refers to the countries where English has traditionally been considered a native, or first, language; i.e. the United Kingdom, the United States, Ireland, Canada, Australia and New Zealand. These varieties are considered to be 'norm providing' (Kachru & Nelson, 1996:78) as they serve as the model for non-native forms of English.
  
- 2) The **outer circle** refers to countries where English is spoken as a second language and where English has essentially acquired an institutionalized form (Kachru, 1982). This comprises countries such as India, Cameroon, Nigeria and Jamaica, to name but a few of the English-speaking territories mentioned in Crystal (2003). For Kachru,
  - 'the main characteristics of such varieties are that (a) they have an extended range of uses in the sociolinguistic context of a nation; (b) they have an extended register and style range; (c) a process of *nativization* of the registers and styles has taken place, both in formal and in contextual terms; and (d) a body of nativized English literature has developed which has formal and contextual

characteristics which mark it as *localized*' (1982:38, emphasis in original).

- 3) The **expanding circle** comprises all the countries where English is spoken as a foreign language and where the English used is a 'performance variety'. Kachru gives the examples of Iran and Japan, where English is taught at schools but does not have societal or cultural function.

'The performance varieties of English have a highly restricted functional range in specific contexts; for example, those of tourism, commerce, and other international transactions' (1982:38).

These varieties, unlike those of the inner and expanding circle, do not have an 'institutionalized status.' Instead '[i]dentificational modifiers such as *Japanese* English or *Iranian* English are indicative of geographical or national performance characteristics' (Kachru, 1982:38).

For Kachru, the inner circle varieties are clearly defined and it is quite simple to determine which countries belong to this circle, but the barrier between the outer and expanding circles is less marked.

'An institutionalized variety always starts as a performance variety, with various characteristics slowly giving it a different status. The main characteristics of an institutionalized variety seem to be (a) the length of time in use; (b) the extension in use; (c) the emotional attachment of L2 users with the variety; (d) functional importance, and (e) sociolinguistic status. In the development of non-native models, two processes work simultaneously: the *attitudinal* process and the *linguistic* process' (Kachru, 1982:39).

It is only recently that the spread of English within Europe has been paid substantial attention (Phillipson, 2003, Cenoz and Jessner, 2000) as people have come to realize that there too, English is an important language for communication. In the past fifty years (Hoffman, 2000), the status of English in Europe has changed in many countries from being a foreign language to being a language with economic importance in the country itself.

The growing importance of English is not without consequence to its structure; the more it is used, the more likely it is to acquire new distinctive forms. Indeed, recent research conducted across Europe has come to the conclusion that English is in the process of acquiring a new Pan European form (see Erling, 2004, Seidelhofer, 2001). The project team's hypothesis concerning the possibility that there is focussing in Switzerland is echoed by Ferguson in terms of Europe:

'English is widely used on the European continent as a international language. Frequently conferences are conducted in English (and their proceedings published in English) when only a few of the participants are native speakers. At such conferences the English spoken often shows features at variance with the English of England but shared by the other speakers. Continental meanings of *eventual* and *actual*, continental uses of tenses, calques or French formulas of conference procedure, various details of pronunciation and dozens of other features mark the English as an emerging continental norm. Native speakers of English attending the conference may find themselves using some of these features as the verbal interaction takes place.' (Ferguson, in Kachru 1982:x).

As noted in chapter 1, Switzerland may somewhat resemble such a conference. English is used as a lingua franca within Switzerland, just as it is in other European countries. An important difference, however, is that English is used far more intranationally in Switzerland than in most or all of the other European nations as was demonstrated by the analysis of the census results.

The question therefore arises: into which of Kachru's circles should Switzerland be placed? Kachru's (1982) and McArthur's (1998) classifications of the native and non-native varieties of English across the world have quite naturally tended to put Switzerland in the third category (even if only by virtue of its not being included in discussions of the first two categories). This makes it part of the expanding circle and an English-as-a-foreign language country. Following the professed language policies within the country and the way English is taught at school, this seems reasonable. However, the growing importance of English in Switzerland since 1960 and especially the way in which it is used by Swiss speakers to each other *within* the

country lends it characteristics of an English as a Second Language rather than of an English as foreign language.

As such, Switzerland may be on the cusp of becoming an ESL country and is an ideal site for the study of the stages which a language may progressively go through on its way to becoming an institutionalized variety. This transition from an expanding circle to an outer circle variety of English, from a performance to an institutionalised variety, from a diffuse to a focussed form has occurred before in the cases of Indian and Nigerian English, for example. Recall, however, that although focussed forms have been studied in detail, the process of focussing as it happens has not.

## 2.4 Aim 1: focussing in Switzerland

To what extent can this previous research give credence to the hypothesis that Switzerland is in a situation where English could be focussing? The following points summarize the main insights drawn from the discussion above.

- 1) **Tertiary hybridization:** The French, German and Italian native speakers of Switzerland are using English with one another, rather than with native speakers of English, which is exactly the type of situation that Whinnom stated would be likely to be involved in tertiary hybridization.
- 2) **New dialect formation:** When English was first used as a lingua franca in Switzerland, each of the three linguistic groups would have used forms influenced by their native language. So there would have been French-influenced English, German-influenced English and Italian-influenced English. If the English in Switzerland has focussed, then these forms, or varieties, would have combined together in a similar fashion to the native varieties in Trudgill (2004) and Kerswill's (2002) research.
- 3) **World Englishes:** The use of English in Switzerland means that it increasingly should be considered to be a second rather than a foreign

language country. This is a vital point because ESL countries are sources of new and distinct varieties of English.

These points underline that the situation of English in Switzerland presents many aspects that allow focussing to occur; one of the aims of this thesis is then to establish whether this is in fact the case. The second half of this chapter will discuss some of the methodology which can be used to establish whether the English in Switzerland is focussed. This second half will also introduce the second aim of this thesis, which is to establish how inherent variability is used by non-native speakers.

## 2.5 Comparing linguistic varieties

There are a number of different ways that English, as used by non-native speakers in Switzerland, can be evaluated in terms of whether there is focussing or not. The most obvious way to consider this is to simply compare features across potentially different dialect groups to determine whether they share similar traits.

Although some features, such as final stop deletion or conjugation regularization, are thought to be *vernacular universals of English* (Chambers, 1995:242-43) and have been found to have developed independently in a number of different English dialects, many features exist in different dialects only through direct contact or transmission. The past participle form of *got* found in some varieties of American English (i.e. *gotten*) finds its origin in earlier varieties of British English and is still used in relic areas of the British Isles (in Buckie in Scotland for example, Smith, 2000:175).

In a similar way, comparing the English spoken by the three Swiss groups can help determine whether the English by them used is a variety separate from Standard English.<sup>14</sup> It can also provide the researcher with valuable clues to determine whether the varieties of English as used by the French, German and Italian speakers have points in common and may in fact be one same variety.

---

<sup>14</sup> In this thesis, I will use Standard English as the version of native English which non-native speakers are most likely to be aspiring to, at least in terms of teaching.

However, such a surface comparison of the features does not always provide sufficient evidence for postulating the existence of a single language variety, as similarity on the surface may disguise underlying differences. This is especially true for features which are variable; that is to say features which demonstrate two (or more) different forms (called variants) that convey similar meaning but which nevertheless are governed by variable rules which dictate when one variant is more likely to occur (e.g. Chambers, 1995:17). Although two language varieties may appear to share a common feature, only an in-depth analysis of how the feature is used in the two varieties can reveal whether it is used in the same way; ‘variationist theorists seek to discover usage patterns in the relative frequency of occurrence or co-occurrence of structures, rather than simply in the existence or grammaticality of those structures’ (Poplack and Tagliamonte, 2001:7).

Variation theory is able to examine the underlying patterns of a feature and it is precisely these patterns that can help us uncover features shared across varieties. That is why I have chosen to use this methodology, especially in terms of comparative sociolinguistics (Tagliamonte, 2002), to determine whether the English of the three Swiss groups is in the process of merging.

This strand of linguistic research, an offshoot of language variation and change, will provide the main driving force behind my analyses and will be presented in detail in the sections below; section 2.6 will present language variation and change, while 2.7 will focus on comparative sociolinguistics more specifically. Because this thesis focuses on variation in non-native speakers of English, it must present some aspects of second language acquisition as well, especially in terms of the studies which have already focused on the interaction of second language acquisition and language variation and change. These aspects will be considered in sections 2.8 and 2.9.

## **2.6 Language Variation and Change**

Before discussing comparative sociolinguistics, I will briefly present some of the main aspects of Language Variation and Change (LVC) and more specifically of variation theory. LVC examines the microlinguistic side of sociolinguistics considering how specific linguistic aspects are affected by society at large. It focuses

on structures, phonological or morphosyntactic, which ‘convey very different social meanings’ (Chambers, 2002:4) although their grammatical meaning is functionally identical (examples 1 and 2 from Chambers, 2002:3).

(1) Adonis saw himself in the mirror

(2) Adonis seen hisself in the mirror

The methodology was first developed by William Labov (1972), and one of the main assumptions is that there is little free variation in language. Rather the main part of variability found is due to either social or internal factors whose relative importance can only be established through an in-depth analysis of the feature (Labov, 1994, 2001). To understand ‘change in progress, we must separate the variation due to change from the variation due to social factors like sex, social class and social networks, and ethnicity, and from the variation due to internal factors like sentence stress, segmental environment, word order and phrase structure’ (Labov, 1994:26).

LVC is most important for this thesis in terms of the methods it uses to come to its conclusions. For this reason, the two sub-sections below will focus on the main terminology used in LVC (2.6.1) and multivariate analysis (2.6.2).

### **2.6.1 Methodology of LVC**

In considering structures that vary, LVC examines features that have two or more variants. In a sociolinguistic analysis, all instances, or tokens, of the variants which are deemed to be part of the variable context are extracted from the data so that the various factors (or factor groups) which might affect the selection of one variant over another can then be considered.

For example, Labov examined realisations of /r/ in non-prevocalic position in New York City (Labov, 1966:64). He found two variants, the occurrence of /r/ (example 3) or the omission of it (example 4).

(3) car - /kɑr/

(4) car - /kɑ:/



By examining all the cases of both variants and considering the external and internal factors which could be thought to influence the variability, Labov was able to determine how the variation was structured.

Factors which condition variability have been found to be either external or internal. External factors are essentially the factors which are separate from a given feature, i.e. the gender or the social class of the person who uttered the variable examined or the formality of the situation in which the variant occurs. Internal factors are associated much more closely to the feature itself as they are part of the language's mechanisms (i.e. phonetic, phonological, morpho-syntactic, pragmatic, etc.) and as such vary considerably from feature to feature (Labov, 2001). In the case of the variation between past tense strong verb forms in found examples 1 and 2 above, an internal factor might be the lexical verb.

In Labov's 1966 research, he found that the external factors of social class, gender and speech styles were very closely linked to the use of the variants. The variation found in /r/ vs. zero was not random but instead was conditioned by the speaker and also by the formality of the situation. In New York, the /r/ variant was the more prestigious one and Labov found that it was used more frequently by the higher social classes and by women. He also found, however, that the rates at which the different variants were used were influenced by speech style. All of the speakers used more of the zero variant in the least formal style (casual speech) than in the more formal styles (reading passages or minimal pairs). The considerable variation found in language does not tend to be free variation but rather factors interact together to form an 'ordered heterogeneity' (Weinreich, Labov and Herzog, 1968:100-101).

By assessing how different factors affect variability, Language Variation and Change researchers are able to establish which factors are conditioning the choice of one variant over another. However, as various factors have been found to interact quite considerably with one another (gender and social class for instance, Milroy, 1987), multivariate analysis is often used to determine the relative importance of the different factors.

### 2.6.2 Multivariate analysis

From Labov (1966) onwards, variationists realized that language is influenced by multidimensional factors and needed to find a way to model this. In many cases, studying factors individually does not allow the linguist to grasp the overall picture of variation, as factors are often impossible to disambiguate through normal distributional means. Multivariate analysis allows researchers to consider a number of different factor groups together and determine which ones are significant and are influencing the selection of the variants when all are considered simultaneously (Bayley, 2002:124, Young and Bayley, 1996:253).

Although multiple regression is a common statistical technique, ‘the programs known as VARBRUL<sup>15</sup> have been used most extensively in sociolinguistics because they have been deliberately designed to handle the kind of data obtained in studies of variation’ (Bayley, 2002:124). Multivariate analysis allows researchers to obtain the variable rules of a feature, these are a ‘mathematical construct mirroring the systematic choice mechanism of language’ (Tagliamonte, 2006:266). Speakers make systematic choices when they use language and because of the ‘systematicity of the process the relative frequency of selection can be predicted’ (Tagliamonte, 2006:130).

Multivariate analysis operates by running a ‘step-wise procedure of [...] multiple regression’ (Tagliamonte, 2006:140).<sup>16</sup> The first step in this process

‘is to find the group which makes the most significant change to the model when it is added or subtracted from the rest. All factor groups are tested, in order to determine which one increases the likelihood most significantly’ (Tagliamonte, 2006:140).

After finding the most significant group,

‘the program retains the most significant groups and tries to add a second group, which increases the likelihood as significantly as possible.

---

<sup>15</sup> These programs are Varbrul and Goldvarb, see Young and Bayley (1996) for an in-depth discussion of how to use the two programmes.

<sup>16</sup> Please see Tagliamonte (2006:128-157) and Young and Bayley (1996) for a more detailed presentation of the processes involved in multivariate analysis.

It continues in this way until no further additions result in a statistically significant improvement' (Tagliamonte, 2006:140).

The step-down analysis repeats this process but in reverse, 'calculating the likelihood of the model when all the factor groups are included in the regression simultaneously' (Tagliamonte, 2006:143) and then discarding the groups 'whose loss least significantly reduces the likelihood (using the chi-square test)' (Tagliamonte, 2006:143).<sup>17</sup>

The results of a multivariate analysis are generally presented in table form, as in table 2.2 below.<sup>18</sup> This table provides the results for a multivariate analysis examining the probability of *wh-* relative clause variants (i.e. *who* or *which*) being selected over *that* or zero variants by English native speakers. This analysis considered three factors: syntax and animacy of relative clause (examples 5-7), adjacency (examples 8-10) and definiteness (examples 11-12).<sup>19</sup>

- (5) SUBJECT ANIMATE: I've found someone **who** is willing to do the homepage (h, German, e-mail)
- (6) SUBJECT INANIMATE: something **that** will make things easier (b, Italian, e-mail)
- (7) OBJECT: I need deans, professors, doctors to whom you have a good contact (c, Italian, e-mail)
- (8) ADJACENT: If anybody knows somebody **who** is keen in design. (b, Italian, e-mail)
- (9) SEPARATED BY RELATIVE CLAUSE: for those who are arriving on the 3. or 4. march in malta (or **who** are staying longer) (h, German, e-mail)
- (10) SEPARATED BY 'ELSE': She also adviced to have a nurse or somebody working with older people **that** can bring us (b, Italian, e-mail)
- (11) DEFINITE: here is the address of the man **who** was in charge of organising (b, Italian, e-mail)

<sup>17</sup> Unless there is a high level of interaction between two of the factor groups, the step-up and step-down runs are identical.

<sup>18</sup> This table provides one of the multivariate analysis runs presented in table 6.22 in the chapter on relative pronouns.

<sup>19</sup> The examples and the factor groups will be discussed more fully in chapter 6.

- (12) INDEFINITE: How we must speak with someone *who* wants to stop smoking (f, French, e-mail)

Table 2.2: Multivariate analysis considering the probability of occurrence of *wh*- relative pronoun forms in English

	Factor Weight (FW)	%	Total N
Input	.41	47	266
<b>Syntax and animacy of relative clause</b>			
Subject animate	<b>.90</b>	87	126
Subject inanimate	<b>.37</b>	28	46
Object	<b>.06</b>	4	94
<i>Range</i>	<i>.84</i>		
<b>Adjacency</b>			
Adjacent	[.50]	46	239
Non-adjacent	[.51]	56	27
<i>Range</i>			
<b>Definiteness</b>			
Definite article	[.47]	41	146
Indefinite article	[.54]	55	120
<i>Range</i>			

The input and the scores of each individual factor in a factor group can range from 0 to 1; values above .50 are said to favour the use of the factor and those which are below .50 are said to disfavour it. The ‘relative position of the factors weights, vis-à-vis each other’ is what is ‘the relevant criterion for interpreting the results,’ however (Tagliamonte, 2006:145).

Input (or corrected mean) is the ‘overall indication of the strength of the rule’ (Young and Bayley, 1996:270). In our case, an input of 0.41 confirms that overall *wh*- forms occur less often than the other variants, which we knew from its overall distribution (47%). In the case of the syntax and animacy of the relative clause, the analysis revealed that while subject animate relative clauses highly favour *wh*- forms (at .90), object relative clauses disfavour them (at .06). This means that the *wh*- variants are very likely to be found used in subject relative clauses but not in object relative clauses. The relative contribution the different factor groups can be determined by considering their *range*; which involves subtracting the factor with the lowest score in

a factor group from the factor with the highest score. Factor groups with high ranges are those which are strongest for the occurrence of the variant under consideration.

The values of some factor groups have been given in square brackets. This is the case of adjacency and definiteness in the table above. This signals that the multivariate analysis did not find that they provided a significant contribution to the occurrence of the variant.<sup>20</sup> Some researchers simply do not include the figures for these factor groups; this thesis will, however, because in comparing multivariate analyses of different varieties, they can be used to uncover similarities between the varieties (Poplack and Tagliamonte, 2001:92). Section 2.7 below will further detail how multivariate analysis can be used to establish the extent to which varieties are related.

More detailed presentation of the results of multivariate analyses will be given in the chapters where it is used. It was important to at least introduce it here, however, because it is an intrinsic part of the methodology of comparative sociolinguistics presented below.

## 2.7 Comparative Sociolinguistics

Comparative sociolinguistics has been used to determine relationships between varieties of the same language. Because ‘[v]ariation theory assumes that the same linguistic *function* may at times be realized in different *forms*’ (Poplack and Tagliamonte, 2001:8, emphasis in text), we can compare features across varieties to determine if they share the same underlying patterns and constraints even in situations when the surface patterns appear to be very different.

This is because ‘[d]ifferences in overall rate of variant occurrence may be due to any number of (non-linguistic) factors’ (Poplack and Tagliamonte, 2001:92) but crucially ‘the conditioning of variability (i.e. the configuration of factors affecting the occurrence of the variant forms), as well as the direction of their effects, are deeper constraints, remaining constant regardless of the extra-linguistic circumstances’ (Poplack and Tagliamonte, 2001:92). If two varieties present the same conditioning

---

<sup>20</sup> The values for this are obtained from the first stepping down run.

and direction of effects in a multivariate analysis then it is highly likely they come from the same source variety.

The methodology of comparative sociolinguistics has been used most impressively by researchers attempting to uncover the sources of African American Vernacular English (AAVE). There is much contention as to where some of the distinctive features of AAVE come from. Some researchers argue that that early African American English was in fact a pidgin or a creole (see Baugh, 1983, Dillard, 1972, Rickford, 1999). Others, however, argue that although the features are not found in mainstream American or British English, they may well come from non-standard varieties of English which were spoken by the slave owners and overseers (Poplack and Tagliamonte, 2001, Schneider and Montgomery, 2001). The former are known as proponents of the Creolist hypothesis while the latter have been said to focus on an Anglicist hypothesis. The difficulties in determining what the actual sources of AAVE are lie in the distance in time and the fact that there is relatively little known either about the English the early American colonists spoke or what varieties the slaves spoke (Poplack and Tagliamonte, 2001:2).

While those who believe in the Creolist hypothesis tend to compare African American Vernacular English with English-based creoles (Rickford, 1999), those who follow the Anglicist hypothesis have instead attempted to uncover earlier forms of both mainstream American English and African-American English, using for example the Ex-Slave Recordings compiled in the 1920s (Bailey, Maynor and Cukor-Avila, 1991) or written letters recovered from the American Civil war era (Schneider and Montgomery, 2001).

The research of Poplack and Tagliamonte (2001 as well as a considerable amount of earlier work) has also attempted to uncover what earlier forms of African-American Vernacular English might have been like by examining the varieties of African American English spoken in relic areas. These areas had little contact with mainstream varieties of English which means that many earlier features have been maintained. Two areas were found to be suitable for this type of analysis, Samaná in the Dominican Republic and the communities of Guysborough Enclave and North Preston in Nova Scotia in Canada.

Speakers from both these areas are descended from freed slaves who left the United States, in the 1820s in the case of the Samaná inhabitants and the late eighteenth century for most of the Nova Scotia inhabitants. Both areas were and to some extent continue to be relatively isolated; Samaná is surrounded by Spanish speaking communities, while the communities in Nova Scotia had little interaction with the white population around them (Poplack and Tagliamonte, 2001).<sup>21</sup>

Most of the research in these areas has focused on comparing the results and figures for these varieties with Creole varieties of English, mainstream varieties of English, and non-standard and relic varieties of English still spoken in the United Kingdom. The features considered have been for the most part morphosyntactic including the tense system, auxiliary inversion, copula deletion and relativization strategies (see Poplack 2000 and Poplack and Tagliamonte 2001 for many of the studies themselves). As Poplack and Tagliamonte see it

‘[t]he operating principle of the comparative method is as follows: if two or more languages share a feature which is unlikely to have developed by accident, borrowing, or as the results of universals, then that feature is assumed to have arisen only once and to have been transmitted from a common ancestral source. The goal is to recover as much as possible of that source from a comparison of the descendants, as well as to determine the changes that have taken place in each’ (Poplack and Tagliamonte, 2001:96).

Their argument why underlying similarity cannot be due merely to chance rests on the results of multivariate analyses. If the analyses of two varieties can be shown to be similar, then the two varieties may have the same variable rules.

If multivariate analysis can prove that a feature has similar variable rules in two varieties, then this feature is judged to have come from a common source variety. This shows that the two varieties are related because the similarity in variable patterns cannot be due merely to chance. Although surface similarity can be misleading,

---

<sup>21</sup> This is especially clear when comparing the results for various features of Guysborough Enclave with those of Guysborough Village, the white village nearby.

similar underlying factors and hierarchies are thought only to occur in related varieties.

By comparing multivariate analyses of the same feature across several varieties, Poplack and Tagliamonte have been able to determine what the most likely links among the groups are. In order to compare the analyses of the various groups, the methodology focuses on three aspects. In their own words:

‘Three lines of evidence deriving from the variable rule analysis may be adduced in interpreting its results: (statistical) significance of effect (at the .05 level), *magnitude* of effect, as determined by the *range* between the highest and lowest factor weight in a factor group, and *hierarchy of constraints*, or ordering of factor weights within a factor group’ (Poplack and Tagliamonte, 2001:92).

The statistical significance is the least reliable of these aspects, as it is the most susceptible to outside influence (i.e. differences in the number of tokens, for example). Poplack and Tagliamonte (2001:93) explain that ‘when a variable is affected in the same way by the same set of factors in several studies, a study with a larger number of tokens will tend to detect more of these factors as statistically significant than one with fewer tokens.’ Because this thesis, like Poplack and Tagliamonte’s research, compares several varieties at once, statistical significance will be considered, but always alongside range estimates and hierarchies of constraints which are affected less by token numbers (Poplack and Tagliamonte, 2001:93).

These two aspects, magnitude of effect and hierarchy of constraints (or constraint ranking) need to be presented at greater depth, because this thesis will use them to present its claims and findings. Both of these lines of evidence are essential in order to properly understand how previous comparative sociolinguists have interpreted their results.

‘The constraint hierarchy subdivides the “sometimes” class into a more refined set of categories, yielding the detailed structure of the relationship between variant and context, or the “grammar” underlying the variable surface manifestations. Even more important for the comparative endeavor



we undertake here, the complex picture offered by the constraint hierarchies serves as an important check in controlling for language universal, and assessing the relationship and provenance of forms' (Poplack and Tagliamonte, 2001:96).

Under the appropriate circumstances, varieties can be shown to come from the same source variety, if the ranges and constraint rankings can be found to be shared across varieties. For example, Poplack and Tagliamonte (2001) have shown that the constraints governing the use of verbal *-s* (i.e the use of a verb form with *-s* in third person singular and plural subjects: *the men eats*) are similar in Samaná and in Devon (Godfrey and Tagliamonte, 1999) but that both are different from Creole varieties of English. They have demonstrated this by showing that Devon and early African American English varieties had the same constraints in terms of subject type, adjacency, phonological context and aspectual effects (Poplack and Tagliamonte, 2001:186ff) and that these results did not match what had been reported for Creoles (Poplack and Tagliamonte, 2001:198).

Using the comparative methodology, Poplack and Tagliamonte have demonstrated a number of times that the patterns found in the relic varieties of AAVE they examined are far closer to the patterns found in other non-mainstream varieties of English than to the patterns found in Creole varieties. The reason that AAVE appears to be radically different from modern day Standard American English is due to the fact that at its roots were a number of different varieties of English which were not part of the makeup of American English as we know it today.<sup>22</sup>

Applying this methodology to non-native variation will be most useful in determining whether a single variety of English is spoken in Switzerland. If speakers from all three backgrounds can be shown to have the same constraint rankings and magnitude of effect for a given variable, then we can establish that they are using the same variety

---

<sup>22</sup> Not all researchers in the field agree with their findings (Creolists in particular), especially in the case of analyses where the data was unevenly distributed making several interpretations possible (as in a few of the studies in Poplack, 2000) (Rickford, 2005). The consensus, however, is that the methodology of itself is perfectly sound.

of English. Moreover, by comparing their results to those of native speakers, we will be able to determine whether their patterns are unique to Switzerland.

The methodology used in comparative sociolinguistics will be slightly modified for the purposes of the current research given that we are dealing with native and non-native groups. The variation of a set of selected features will be analyzed, using multivariate analysis (when possible); and then the variation patterns and hierarchies found in the speech of native speakers will be compared to those of Swiss non-native speakers (compare to Regan 1996, and Rehner, Mougeon and Nadasdi, 2003 and see section 2.9 for more discussion of non-native variation). This will allow us to ascertain whether the non-native patterns are due to the acquisition of native speaker patterns or whether they are due to second-language learning processes. If the non-native groups share the patterns of the native speakers then the variability can be judged to have been acquired, as following Poplack and Tagliamonte's methodology the similarities are unlikely to be due to chance.

The next point in this process will involve comparing the three language groups (French, German and Italian) with each other to see how their patterns compare. If they do not share the same patterns then there is no focussing, and if they do, and these patterns are different from native patterns, then I will argue that there is focussing.

The decision to use comparative sociolinguistics entails that it is necessary to choose for study those features which are used variably (Young and Bayley, 1996) and will thus present rates and determining factors. This restricts the type of features examined somewhat, but, in actual fact, even features that have traditionally been studied in Second Language Acquisition studies are variables to some extent (in that there is a Standard English variant and a non-native one). The variability considered can either result from interference from a Swiss language, or the presence of variability in the target language. Comparative analysis will enable me to disentangle these multiple possibilities.

The methodology of comparative sociolinguistics has hitherto been used primarily on native varieties of English, either those with uninterrupted paths of transmission or

creoles. Section 2.8 below will present the studies which have used the methodology of LVC as well as comparative sociolinguistics in terms of non-native variability. This section will also show how my own research in this thesis is more extensive than previous studies of second language variation in its aims and its use of comparative sociolinguistics because it compares several non-native varieties to each other. Section 2.9 will briefly discuss aspects of Second Language Acquisition (SLA) and demonstrate how the focus of SLA is considerably different from that of Second Language Variation.

## 2.8 Second Language Variation

Although research into the field of language variation and change is quite vast (see Chamber, Trudgill and Schilling-Estes' *Handbook of Language Variation and Change* (2002) for an idea of the breath of variation studies), there appears to have been a dearth of variation studies dealing with 'aspects of the target language where native speakers display linguistic variation' in the output of non-native speakers (Mougeon and Rehner, 2001). Even very useful collections such as Preston and Bayley's *Second Language Acquisition and Linguistic Variation* (1996) contain numerous papers that deal almost exclusively with variation due to interference of the source language rather than the acquisition of native-like variation.<sup>23</sup>

In the past few years, however, a number of researchers have concentrated their efforts on Second Language Variation (SLV)<sup>24</sup> studies, and panels at important sociolinguistic conferences (Sociolinguistic Symposium 15 in 2004 and NWAV 33 in 2004) have focused solely on discussion of variation in high level bilinguals. This research has chosen to distinguish itself from 'the bulk of previous research in Second Language Acquisition (SLA) [which] focused on aspects of the target language where native speakers display **invariant** language usage (i.e., use only one linguistic element to convey a given notion)' (Mougeon and Rehner, 2001:398, emphasis mine).

---

<sup>23</sup> Regan's paper on the acquisition of *ne* by Irish university students of French is a notable exception.

<sup>24</sup> This abbreviation is used exclusively in this thesis and does not have the currency of SLA or LVC, although it does present a useful counterpoint.

Indeed this “new strand of research” (Mougeon and Rehner, 2001:398) is quite different from traditional research into second language acquisition in that it:

“includes not only factors that have been examined by mainstream SLA research, but also those that have been found to be correlated with L1 variation in sociolinguistic research” (Mougeon and Rehner, 2001:399).

This type of research looks for proof that L2 learners can show ‘the same kind of sociolinguistic ability in using the variants as do L1 speakers (i.e., ability to observe the linguistic and extralinguistic constraints that have an impact on variant choice)’ (Mougeon and Rehner, 2001:399).

If non-native speakers are to be considered to have fully mastered the target language, they must show they have acquired the syntactic and phonological aspects of the non-native language. The non-native speakers will also have to display that they have acquired the variable rules of native speakers, both for features where the variation is stylistically motivated and those where the variation is internally constrained. These variable rules belong to native speakers’ communicative competence (Hymes, 1972:281) and are an intrinsic part of the mastery of one’s own language. Using the methodology of comparative sociolinguistics will allow us to test how non-native speakers deal with this inherent variability and to determine the types of variable constraints that they are able to acquire.

A number of the SLV studies have shown that sociolinguistic competence is not as easy to acquire as other features of the target language (see for example, Dewaele, 2004, Dewaele and Regan, 2002, Regan, 1996, Rehner, Mougeon and Nadadi, 2003). It has been found “that immersion students learn an academic register of the L2 but not its vernacular” (Lyster, 1996:167). This is because “whereas there exist numerous dictionaries and reference grammars to support the teaching of lexis and syntax, there are no such reference books to support the teaching of sociolinguistic variation” (Lyster, 1996:167). It is not surprising that there are no reference books which teach sociolinguistic variation, as for the most part even native speakers are not aware of the variable rules they use everyday. The fact that the application of these variable rules is almost totally subconscious for native speakers means that they can be more difficult

for non-native speakers to notice and acquire. The difficulties experienced by the students in these studies are not restricted to external constraints but to internal factors as well (as demonstrated in Regan, 1996).

Rehner, Mougeon and Nadasdi (2003) examined non-native French speakers in immersion classrooms in Canada, while Nagy, Blondeau and Auger (2003) and Sankoff, Nagy, Blondeau, Fonollosa and Gagnon (1997) studied other speakers of non-native French in Canada. An important part of Regan's current work (Regan, 1996, Dewaele and Regan, 2002) focuses on French non-native speakers in Europe. Bayley has conducted a considerable amount of research on the variation patterns of Chinese speakers, either learning Chinese as a heritage language in the U.S., or learning other languages abroad (Bayley, 1995, Langman and Bayley, 2001).

These few examples demonstrate that a significant proportion of the research conducted into second language variation has been concerned with non-native variability in languages other than English. While findings for non-native variation in French and Chinese can help us predict what might be found in non-native speakers of English, they cannot tell us precisely how non-native speakers deal with inherent English variability. Thus far there has been very little research into non-native English variability. The present thesis aims to redress this by considering how three different linguistic groups deal with native variable patterns.

## **2.9 Second Language Acquisition**

The Second Language Variation research presented above diverges from general research in Second Language Acquisition in a number of ways; the differences which are the most important for the present study have to do with the type of non-native speakers examined and the type of non-native variation studied. Because the main focus of this work is not the acquisition of a foreign language in general but on the acquisition of variable features, this section will focus primarily on exemplifying precisely how SLV and SLA differ and only briefly discuss a few relevant points of SLA research.

### **Interlanguage vs. high level competence**

Second language acquisition studies focus on the ways people acquire a language that is not their mother tongue (Ellis 1985, 1994, Tarone, 1988). For this reason, many studies of SLA focus on the initial stages of acquisition and the processes that are found in speakers as they first start to acquire a new language. An important concept for the study of the initial stages of language acquisition is **interlanguage** which is seen as an intermediary stage of the L2 and which contains a considerable amount of interference of the L1 (Tarone, 1988).

This is problematic in terms of the Swiss situation, as interlanguage implies a temporary stage of language and speakers of English in Switzerland, at least those who belong to the group considered in this research, are not in the process of learning English and should not be considered to be using an intermediary type of English.

A distinction which needs to be made, but is not always made explicit in the literature of second language acquisition, is that there are not only second/foreign language learners but second/foreign language speakers as well. The latter should be taken to mean speakers whose second language use has gone far beyond the initial acquisition stages and whose language can no longer be considered an interlanguage. These speakers may have a number of non-native features or possibly idiosyncratic features in their speech, but their language is stable (but not invariable) and fluent. Both the speakers examined in the present study and those considered by other Second Language Variation researchers fall into this category of high-level second language speakers.

### **Types of variation**

Because much research in SLA has focused on interlanguage varieties, the types of variation considered are also rather different than that studied by SLV researchers. Interlanguage contains a lot more variability than is found in more fixed forms of speech and the general consensus of second language acquisition researchers is that it is largely of the *free variability* type.

‘When learners first internalize new linguistic items, they do not know precisely what functions they realize in the target language. The result is free variability’ (Ellis, 1985:85).

The variation found in learners is non-systematic and the ‘new linguistic items’ are used interchangeably. While this is undoubtedly accurate for interlanguage, the situation is rather different in terms of the language of more proficient non-native speakers. In their case, while there is some free variability, there are also far more cases of variation which is systematically ordered. The assumption that most of the variability found in non-native speakers is non-systematic has as a consequence the fact that when SLA researchers consider variation they tend to examine features that can help reveal the inner workings of interlanguage. These features are generally those which deviate, from native norms. Examples 13 and 14 below provide examples of a feature that researchers of SLA are likely to consider; the overuse of the infinitive form in cases where Standard English has *-ing*.

- (13) For this, I write here the way for helping students who want to stop *to smoke* (f, French, e-mail)
- (14) So, I look forward *to read* your suggestions! (b, Italian, e-mail)

What is neglected in these studies is the type of variation that is neither free nor problematic; the type of variation considered in native speakers as discussed in section 2.6. While it may be difficult to consider such variability in the L2 of speakers acquiring a language, it is not so for high level non-native speakers as much of the initial non-systematic variability will be reduced. The researchers mentioned in section 2.8 have looked precisely at features which show inherent variability in native speech and have conclusively shown that there is much to be uncovered about the processes of second language acquisition when dealing with features associated with sociolinguistic competence in the target language.

Moreover, in some cases, the notion of free variability is taken too far and even features which have been shown by variationists to have variable rules are considered to be free variability:

‘I alternate haphazardly between ‘who’ and ‘that’ as subject relative pronouns in non-restrictive relative clauses.’ (Ellis, 1985:80).

The alternation in the choice of relative pronouns is most definitely a case which does not involve free variability, as has been demonstrated by a number of researchers (Ball, 1996, Guy and Bayley 1995, Tagliamonte, Smith and Lawrence, 2005, see

chapter 6 for a full list) and in the present thesis (see chapter 6). A consequence of this is that features which might help us understand second language acquisition better have been left aside in SLA research.

The present thesis considers the type of features (which deal with systematic native variation) that have been neglected by SLA researchers in order to determine what these can tell us about the acquisition of variability and about second language acquisition in general.

## **2.10 Aim 2: Non-native acquisition of variation**

Recall that the first aim of this thesis is to determine whether the patterns found in the three non-native English speaking groups in Switzerland are the same and whether focussing might be underway in Switzerland. The second aim is to test the hypothesis that non-native speakers are able to acquire the structured heterogeneity of the native English speaking community. This will allow us to determine if native variable patterns can be transmitted intact to non-native speakers.

Studies into second language variation have thus far focused primarily on French, so by studying non-native English variability this thesis will be able to establish whether the conclusions of earlier research holds true in another language. Obviously, when speaking of native English variability, we also have to consider what form of English Swiss speakers are aspiring to, as their target will have a bearing on the results.

Swiss speakers of English have Standard English as a model; English is acquired first of all at school in Switzerland. The grammar books used in Switzerland focus primarily on British English, so although American English may provide a strong linguistic model through the media, it is a British model of English that the students will be trying to match. The features considered in this thesis will then need to be ones which show systematic variation in Standard British English.

Rather than considering the variation in speakers from a single linguistic background as had been the case in a number of previous studies, this study will focus on three



non-native linguistic groups at once. This will allow us to compare these groups to one another as well as the native control group.

The two aims of the study are complementary. Examining what form the variation in a given feature takes will allow us to determine to what extent the three language groups use the same form of English and whether it is comparable to native English. Each feature analysed will then be able to provide us with results which, first of all, will allow us to see if Swiss speakers have managed to acquire the variation patterns of native speakers. Secondly it will allow us to determine whether the pattern is identical or very similar across speakers the three source languages, and whether this might be indicative of focussing.

This will be achieved by using the methodology of comparative sociolinguistics outlined above. Each feature will be examined to assess whether its results point to focussing, free variability/interference or native patterns. Each of these outcomes has a different pattern as I will demonstrate in chapter four, but first the data (and community) which I will use for my analyses needs to be presented.

### 3. Data Sources

#### 3.1 Introduction

##### 3.1.1 Data Collection

The results of the national census (chapter 1 and OFS, 2002, 2005) demonstrated that English is overwhelmingly used in the workplace and not at home.<sup>25</sup> The use of English by Swiss speakers is restricted to specific areas and situations where groups with a mix of French, German and Italian speakers need to communicate. This is linked to its status as a *lingua franca*. This situation is very different from native speakers of English who use it in a wide variety of styles and situations.

Consequently, in order to obtain an accurate assessment of English, it was crucial to collect data in the actual situations where English was being used (for example, in e-mail communication), rather than through interviews. This is not to say that interviews are not valuable in the study of the speech patterns of English found in Switzerland. Rather interviews need to be taken with far more caution than data sources which are further removed from the problem of the observer's paradox, the idea that speakers monitor their speech when they know they are being observed (Labov, 1972).

##### 3.1.2 Observer's paradox and non-native speakers

The observer's paradox must be considered in any form of (socio)linguistic collection of course; as discussed by Labov (1972:209-210), 'the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed: yet we can only obtain these data by systematic observation.' There are ways to lessen the influence of the observer's paradox (i.e. talking about topics which are likely to make the interviewees forget themselves, near death experiences, childhood games, etc.) but it is nonetheless important to accept that there are inherent difficulties in collecting accurate representations of people's natural speech.

---

<sup>25</sup> Although the results in chapter 1 combined English used at home and at work, a 2004 press release from the OFS (OFS, 2004) discussed that the increase in the use of English was in the workplace (15.9% in 1990 to 21.7% in 2000).

In Switzerland, where the language we are examining is not the native language of the interviewees, the effects of the observer will be especially strong, as non-native speakers are far more likely than native speakers to pay more attention to their speech in an interview situation. In fact, a number of studies have demonstrated that the use of certain linguistic features (primarily research which deals with interference of the source language on the learner's speech production) is strongly dependent on the amount of attention language learners are paying to their speech (Tarone, 1988; Eisenstein, 1989). Interviews are not the ideal medium for obtaining unmonitored Swiss English data.

### **3.1.3 Aims of the present chapter**

As discussed in chapter 2, this thesis presents analyses of two speaker groups; non-native Swiss speakers of English and a control group of native English speakers. This chapter will introduce the two distinct corpora used in this thesis and will discuss the importance of ensuring that the types of corpora examined are directly comparable. A large section of the non-native data is comprised of the e-mail correspondence of a Swiss medical students' association, which is why the English native speaker data is also from e-mails. Computer mediated data has been found to be somewhat different from both oral and written registers (Herring, 2001) so we must ensure that the data is comparable across linguistic groups. Moreover, this chapter will discuss the ethnographic insights which can be uncovered from this association about the forces that may be behind the change to English in Switzerland.

## **3.2 Swiss data**

Despite it being a known fact that numerous companies in Switzerland use English as their main language (Watts and Murray, 2001), it was much more difficult to collect data from them than might have been anticipated. Although companies willingly admitted that most of their employees used English on a daily basis, they were not willing for the research project to record employees in a natural work setting, however, as the companies deemed that this could be disruptive to productivity. Some of the companies cited industry secret concerns as well. This meant that although the

project was able to collect some data from speakers working in companies that use English on a regular basis, these speakers were interviewed outside of the work environment, so might not have provided an accurate representation of the English they generally use.

Therefore, the present thesis only considers data that was collected in situations where English was used naturally and spontaneously. This means that only a subsection of the data collected for the entire project corpus has been used. In addition to the data used in this thesis (IFMSA-CH) which is presented in section 3.3 below, the project collected data from three other sources. A panel discussion at a music festival held in Zurich in 2002 (M4Music) provided data from six Swiss speakers. Videotaped conferences from the First Tuesday group in Geneva (First Tuesday) provided data from a further 12 Swiss speakers. Finally interviews of eleven employees at an international bank in Zurich where English was used on a daily basis (bank interviews) were obtained.

The conference and the panel discussion were not considered in this thesis because they were in large part scripted and thus cannot be considered to be unmonitored speech. Moreover, there was not always enough data from the individual speakers to be able to gain an accurate picture of their English use. The interviews were excluded from this analysis because, without unmonitored English data from the speakers, the effects of the observer's paradox could not be judged.

This additional data will be used as a point of comparison, however. In chapter 4, it will be considered to help determine how widespread the plural count form of *information* (i.e. *informations*) is in Switzerland.

The data used for analysis in the present thesis consists of instances of Swiss English collected from meetings, e-mails and through a few interviews of members of a selected group. For the most part, the various speakers' mother tongues, ages and other identifying characteristics are known. The group, a Swiss Medical Students' Association, will be presented in detail in 3.3.

Although examining the output of a single group, the association mentioned above, means that we cannot present a full picture of English use in Switzerland, this thesis can, however, accurately present the linguistic situation of one particular group of Swiss English speakers. An understanding of the processes involved in the use of English will be gained by considering the languages used by this group of speakers and the circumstances in which they are used. These findings can then be used to infer whether similar strategies could have been adopted in other Swiss situations.

### 3.3 IFMSA-Switzerland



#### 3.3.1 The Association

The **International Federation of Medical Students' Associations - Switzerland** (hereafter IFMSA-CH) is an association of students who are studying at the various medical schools in Switzerland at the universities of Lausanne, Geneva, Berne, Zurich and Basel (IFMSA, 2003, IFMSA-CH, 2003).

As described by one of the members of IFMSA-CH, the purpose of the association "is to enable international cooperation in professional training and the achievement of humanitarian ideals" (b, Italian, e-mail).<sup>26</sup> The members of the association help Swiss medical students find placements in the hospitals of other countries and organize the Swiss internships (or clerkships as they are called by the members themselves) of students from abroad.

---

<sup>26</sup> The codes for the examples from the data are as follows: the letter code of the speaker, their native language and the medium the example was delivered in. The linguistic background of the members was provided to me by one of the presidents.

Some of the members also attend general assemblies (called GAs<sup>27</sup>) around the world twice a year; this allows them to meet members of the wider association from different countries to organize new exchanges with them. The worldwide federation, of which IFMSA-CH is a chapter, celebrated its fiftieth year of existence in 2001, and is recognized by the United Nations as a non-governmental, non-profit organization (IFMSA, 2003). IFMSA boasts over 80 member countries from all parts of the world.

As in most associations, the IFMSA-CH members hold various positions; the most important of these are the President, the Secretary and the NEO (National Exchange Officer, who is in charge of supervising all exchanges made). Other members are either Local Exchange Officers (LEOs) or in charge of one of the standing committees in Switzerland. These standing committees focused on health projects and issues not directly covered by the National Exchange Officer. Table 3.1 below, presents the Standing Committees in which Swiss students were involved, as well as the main positions in the association. Although elections were held each year for the positions, in a number of cases the same members kept their roles for several years.<sup>28</sup> Table 3.1 provides information on the university and native language of the holders of the main positions within the association over the four years of data collection, in order to provide a clearer picture of their overall background.<sup>29</sup>

Table 3.1: IFMSA-CH members in the main positions 1999-2002				
	1999	2000	2001	2002
President(s)	(GE <sup>30</sup> ) French	(GE) Italian/ (ZH) German	(GE) Italian/ (ZH) German	(GE) French/ (GE) French
National Exchange Officer (NEO)	(BS) German	(BS) Italian	(BS) Italian	(BS) Italian

<sup>27</sup> This term was often used even by members when speaking their native language; rather than using AG which would abbreviate Assemblée Générale in French, for example.

<sup>28</sup> The secretary, for example, was in fact employed by the association and was not in fact elected.

<sup>29</sup> The speaker codes are not given here, as providing the position and the code would have made it possible for the identity of the members to be revealed and the members of the association had been promised that their privacy would be respected insofar as possible.

<sup>30</sup> The abbreviations used are: GE = Geneva, LS = Lausanne, BE = Bern, ZH = Zurich, BS = Basel

Secretary	(BS) Italian	(BS) Italian	(BS) Italian	(BS) Italian
National Officer on Reproductive Health and AIDS (NORA)	(ZH) German	(ZH) German	(ZH) German	(ZH) German <sup>31</sup>
National Officer on Refugees and Peace (NORP)	(GE) French	(GE) French	(BE) German	(BE) German
National Officer on Research Exchanges (NORE)	(GE) French	(GE) French	(GE) French <sup>32</sup>	(GE) French
National Officer on Medical Education (NOME)	n/a	n/a	n/a	(LS) Italian/ (ZH) German

At the time of data collection (throughout 2001 and early 2002) there were approximately 80 members within the Swiss branch of the association.<sup>33</sup> All but one of the members of the association were students and were in their twenties. The one non-student (and older) member was the secretary who had been hired by the association. As can be seen in Table 3.1, the presidency of the association was quite often shared between two members, so that, from 2000 to 2001, the presidents were a member from the University of Geneva and a member from the University of Zurich, and in 2002, the presidency was held by two members from the University of Geneva.

This association is particularly relevant in terms of a project researching possible focussing in Switzerland because in many ways it represents a *community of practice* (or more generally a social network) (Eckert, 2000, Meyerhoff, 2002). Its members ‘get together in order to engage in their shared practices’ (Meyerhoff, 2002:527); moreover, they ‘share some *jointly negotiated enterprise*’ (Meyerhoff, 2002:528, emphasis in original) which in this case is their helping other medical students study abroad. A third criterion for belonging to a community of practice is that the members

<sup>31</sup> The position is occupied by a different member than the previous years

<sup>32</sup> The position is occupied by a different member than the previous years

<sup>33</sup> This was the number of listed members on the mailing list website.

have a shared repertoire; ‘these resources (linguistic or otherwise) are the cumulative result of internal negotiations’ (Meyerhoff, 2002:528). Not only do the members of IFMSA feel they use their own particular brand of English, but they have also made a conscious choice to use English over other languages. This will be examined further in section 3.3.8.

The association provides an obvious case of a dense, multiplex network – exactly the type of situation that Le Page and Tabouret-Keller (1985:182) discussed in terms of focussing. As was discussed in chapter 2, Le Page and Tabouret-Keller (1985:5) consider ‘tightly knit and closely-interactive communities’ to be intrinsically linked to focussing.

The reason social networks and communities of practice are important linguistically is because it has been found that ‘[n]etworks constituted chiefly of strong (dense and multiplex) ties support localized linguistic norms, resisting pressures to adopt competing external norms’ (Milroy, 2002:550). Moreover, it has been found that ‘the adoption of a way of speaking, like a way of dressing, no doubt requires access and entitlement to adopt the style of a particular group’ (Eckert, 2000: 211). As demonstrated in Milroy (1987) social networks are very important for the formation of focussed varieties, and the denser a network is, the more likely it is that the language used by the whole group will be regularized.

### **3.3.2 Languages Used in the Association**

As could be expected from students in a country such as Switzerland, members come from a variety of linguistic backgrounds; French, German and Italian speakers are all represented in the association. The diverse linguistic backgrounds of the members of this Swiss association meant that they had to decide which language or languages to use when communicating with each other.

The universities which the members attend are either German-speaking or French-speaking, as the Ticino (TI), the one Italian-speaking canton, does not have its own medical school at present. For the most part, the members have chosen to go to a medical school in a university in which courses are taught in their native language;



thus the French speakers go to the French-speaking universities (Lausanne and Geneva) and the German speakers choose the German-speaking ones. There are Italian speakers at both French- and German-speaking universities.<sup>34</sup> Although these facts are interesting in terms of the Swiss languages, they will not impact on the analysis of English as the amount of English used is the same regardless of whether the students were at a French or a German-speaking university.

Despite the fact that all the students spoke either French or German as their university language, at the time of data collection, English was the main language used by members of the association. This was the case both for face-to-face interaction, at the twice-annual national meetings and for written communication, i.e. in the e-mails which were sent to a general mailing list for the whole association.

Note that this thesis focuses on the language(s) used for communication at a national level; considering how English was used as a lingua franca within the association. In terms of language choice on a daily basis, within the Local Exchange Offices (based in each university), for example, the problem of lingua franca does not occur. Locally, members would use their university language; so students in Lausanne or Geneva use French, and those at the German-speaking universities use German or Swiss German. It is only when members from different linguistic regions communicate with one another that the question of lingua franca comes into play.

Although English was the language of choice for intranational communication from 2001 onwards, data collected prior to that time and discussions with association members revealed that this was not always the case. Analyzing e-mails sent by members of this association over a period of three years, starting before this switch to English, will allow us to examine how English gained so much importance within the association. Grasping the reasons why the switch to English might have taken place will be crucial to our understanding of what the English spoken by members of this association is like. Section 3.3.4 will focus on this analysis of the increase of English use over time in the e-mails, providing a real-time (Bailey, 2002) example of some of

---

<sup>34</sup> For this reason, in Table 3.4 below, the Italian speakers are assigned two main languages: Italian, which is their mother tongue, and also the language they use for their studies, referred to as university language (French or German). In the linguistic analyses only native language was considered.

the difficulties encountered by Swiss people when selecting a lingua franca for nationwide use. The sections directly below will present the mailing list on which the e-mails were sent in more detail. These e-mails will form the main part of the data used for the subsequent analyses in this thesis.

Examining the circumstances which led to the change to English within this association are very important as, unlike companies or some other associations, the choice to use English sprang entirely from the members themselves and at every point it was a decision which all the members understood and accepted. English was used in this association because it was the best solution for all the members, not because it was imposed on the members.

The fact that membership to this association was relatively limited meant that we could determine what each member's attitudes about using English were and see how homogenous their English was compared to the rest of the group. Of the 80 members listed on the mailing list, we have data for 48 of them. Through the analysis of metalinguistic comments in the e-mails themselves, as well as through discussions with the members, a view of the members' attitudes regarding the use of English can be obtained.

IFMSA-CH's linguistic situation differentiates it from other IFMSA member countries. In many of these countries, one language would be used for all country internal e-mails and information, and another language, i.e. English, for country external information (anything having to do with the association as a whole, as English is the official language of the worldwide association). Consider, for example, France: all e-mails sent to members in the various French universities would be in French but the French members would use English at the general assemblies with members from other countries.

### **3.3.3 The Mailing List**

The mailing list is IFMSA-CH's main method of intranational communication. In May 1999, IFMSA-CH began using a Yahoo mailing list to ensure that all

interested members received information as quickly as possible and to avoid having to send messages individually.

Members sent messages from their own account to the mailing list address; these messages were then automatically forwarded to all other registered members. The format of messages varied depending on the sender. Some members were able to make use of standard word document features (bold, italics, colours) in their e-mails.

Although the messages were of varying lengths, ranging from just a few words to many pages of reports from various international meetings, most were 200 words or longer and thus provided enough data to analyze speaker patterns accurately. Note that among the most frequent writers, some did not in fact write more than a sentence or two most of the time; whereas some writers did not contribute to the mailing list very often, but when they did, it was always with extensive messages. This is why it is often possible to form a clear picture of the English of members who only sent a handful of e-mails. The following e-mail is an example of the language production of the Swiss speakers.

- (1) Hi guys, you remember at the past meeting in Bern a red map that 'b' gave to 'c', concerning the activities of SCOPH and the presentation of their activity? Well, c [...] has lost it somewhere. Has anybody of the participants of the meeting taken it home? Please, answer quickly! I need that map desperately!
- A lot of THANKS
- c (Italian, e-mail)

Features which mark the e-mail as being written by a non-native speaker of English have been underlined ('map' is most likely a calque from the German 'Mappe' meaning folder).

More messages are exchanged at times when events need to be organized (transport and registration for conferences) than when each local exchange office is working individually, but on the whole the mailing list is very productive. Over a period of three and a half years more than one thousand e-mails were collected.

The value of studying e-mails in relation to language choice is two-fold. First of all, messages can be counted to determine which language is used most often, and which ones were used at the start, but then fell into disuse. Second, quite often members talk about language use on the mailing list, or their language use allows us to infer the reasons for specific uses; this provides metalinguistic commentary on their language choices. These will be discussed in section 3.3.8 which covers the causes of the change to English.

The usefulness of the mailing list in investigating language choice derives from the fact that each individual member is relatively free to decide what language to use, as can be seen in the following comment:

- (2) “I repeat: if you don't feel writing in English, but you feel like sharing something with us, please write in your language, ... it is better that not writing anything!” (b, Italian, e-mail)

### **3.3.4 The Situation at Meetings (and Interviews)**

In addition to collecting e-mails from the IFMSA mailing list, I attended two of their national meetings, and interviewed several of the members about their thoughts on language use in Switzerland. These meetings and interviews provide us with oral data to compare the e-mails with. While language use at the meetings was not influenced by my presence, the fact that the interviews were conducted in English undoubtedly was. However, because I had previously met the people interviewed at one of the national meetings, I was able to establish that there was not a noticeable difference in the level formality of these speakers' English between the meeting and the interviews.

The first meeting, in May 2002, was conducted primarily in English, and when German was spoken, there were requests from the French members to change into another language. There were no requests to change from French when it was used; however, conversations overheard afterwards suggest that French was not always readily understood by some of the German speakers. The main speakers at the meeting (who for the most part were the same as the main e-mailers in the study, 'a',

'b', 'c' and 'h'), were quite comfortable switching from one language to another. In fact, quite a lot of code-switching took place with some of the speakers, with sentences starting in German, switching into French, and ending in English, as in example 3 below.

(3) *Also ich habe die das Spiel organisiert und der*, et et pour ça il y a eu quelques problèmes **because I have now three people to present to you** (b, Italian, meeting)

This is especially interesting, considering that very little code-switching was found in the e-mails (see the analysis below). The first meeting was in Basel, which is the *de facto* headquarters of IFMSA-CH, and was very well represented in terms of languages. About 50 people attended, roughly half of them from the French-speaking universities (Geneva and Lausanne), and half from the German-speaking ones (Zurich, Bern and Basel).

The second meeting (October 2002) was quite different. First of all, it did not take place at a local exchange office, but at a high school in Sion (in the canton of Valais) that one of the members had been able to obtain permission to use for the meeting. Second, fewer members were present, only about twenty in total. Furthermore, the ratio of French to German speakers was highly in favour of the French speakers, about three to one. The outcome of this was that this meeting was conducted primarily in French with the German speakers using German when they wished to participate. The main speakers present at the meeting in Basel had either left the association entirely (having graduated from medical school) or were simply not present at this second meeting. Only the first of the two meetings was transcribed for the purpose of the feature analysis.

The meetings serve to remind us that the overall distribution of a group determines to a significant extent its language choices; in a group with considerably more German speakers than French or Italian speakers, for example, the main language will likely be German. English is only spoken in roughly heterogeneous groups, at least within this association.

We learn from language usage at the two meetings and the analysis of e-mails that

English is used in circumstances when a global understanding is sought. In smaller groups, or in e-mails to specific individuals, English is not as important, as it is easier for one speaker to ask for further explanation or clarification. These findings support the hypothesis that English is used as a lingua franca in Switzerland: it is used in cases when the speakers could not otherwise make themselves understood by speaking their own languages.

As well as the data collected at the meeting in Basel, I was also able to conduct interviews with two students based in Lausanne. It was rather difficult for the students to find time for the interviews so I focused mainly on obtaining data from the French speakers as they were slightly underrepresented in the rest of the corpora. Both interviews lasted roughly an hour and the interviewees were asked general questions about themselves as well as their views on English in Switzerland and in the association.

The transcripts of the meeting and the interviews were verified by two people to ensure that the transcription was accurate. Occasional passages remained incomprehensible and have been marked accordingly in the data. The meetings and interviews were transcribed in a standard form with salient phonetic aspects signaled in the transcript, as my own interest in the transcripts linguistically had to do with syntax rather than with phonology.

### **3.3.5 Data and Methods**

There were two periods of data collection for the e-mail corpus; a first period lasting until June 2002 in which every single e-mail sent on the mailing list was collected, and a second from June 2002 onwards when only e-mails sent in English by Swiss members were collected. Data from both periods is considered in the feature analyses but for the purposes of the analysis below which focuses on how English became the main language on the mailing list only the first period is considered. Furthermore, any additional data in the e-mails in the first period (such as attachments and replies included in the message) has been included in the feature analysis but not in the quantitative analysis in this chapter.

The total number of messages collected for the quantitative study was 996.<sup>35</sup> This includes every message sent between May 1999, when the list was started, and June 2002, a total of 38 consecutive months. Thus, counting the number of messages in each language can show us the relative importance of each language. The second period, comprising only English e-mails provided a further 144 messages. Table 3.2 below presents the distribution of e-mails over the four years of data collection. This table also examines the percentage of messages sent by each linguistic group in each period; it shows us that the distribution of the linguistic groups has changed over the four years. It is most evenly balanced between the linguistic groups in 2000, we will come back to this point in our discussion of the results.

Table 3.2: Total number of e-mails on IFMSA-CH mailing list

	Total number of e-mails	% of e-mails sent by French speakers	% of e-mails sent by German speakers	% of e-mails sent by Italian speakers
1999	64	53	28	19
2000	235	38	26	36
2001	332	16	27	57
2002	249 + 144 English only <sup>36</sup>	34	16	50
Total	1024	30	24	36

### 3.3.6 Results

#### E-mail Classification

In order to calculate the percentages of the use of English versus other languages, the e-mails were divided into four categories: messages in English, messages in French, messages in German, and a category of miscellaneous messages. The miscellaneous category includes messages which do not belong in the research, as they were not written by Swiss speakers; messages which, for some reason, were sent twice; and messages in languages other than the three mentioned above.

<sup>35</sup> This number includes messages which were then later discarded and do not appear in the classification

<sup>36</sup> The 144 English only messages are not included in the analysis considering the percentage of messages sent by each linguistic group.

Although the categories above focus on the language used in the e-mail, some messages could not readily be classified as belonging to a single language; in that they started off in French or German and then switched to English or vice versa. Such switches are not uncommon in situations where people share several common languages and are called code-switching (Poplack, 1980). To ensure that messages containing code-switching were not following different patterns over time than the bulk of the e-mails, they were considered separately in a first instance. Before continuing with the analysis by language, this section will focus on the mixed languages. To consider mixed messages, all messages were sorted into one of three groups:

- 1) monolingual messages<sup>37</sup>
- 2) mixed-dominant messages, where most of the message was in one language with a sentence or two in another language
- 3) mixed-balanced messages, e-mails in which two (or more) languages were roughly equally represented.

It was relatively easy to distinguish mixed-dominant (example 4) from mixed-balanced messages (example 5). Messages such as those that started with 'Ciao!' and went on in another language were considered to be monolingual.

- (4) Dear IFMSA-CH Family, the week-end meeting is approaching, here is the list of people who will hopefully come! (I haven't heard anything from IFMSA-Lausanne,... **'g' and 'q', on veut voir les photos de Malte :-)**

(b, Italian, e-mail, translation: 'g' and 'q', we want to see the pictures from Malta.)

- (5) ... und(et) ich(je) komme(viens) auch(aussi) nach(à) basel(bâle) aber(mais) leider(malheureusement) nur(seulement) am(au) samstag(samedi). ich(i) hoffe(hope), ihr(you) könnt(cân) mir(to me) verzeihen(forgive). 'T'.

---

<sup>37</sup> NB: this comprises all monolingual messages and not only those in English.



(T, German, e-mail, translation:... and I'm coming to Basel as well,  
but unfortunately only for the Saturday. I hope you can forgive me)

### **Monolingual versus Mixed-language Messages**

Table 3.3 examines the overall frequency of these three types of message. We can see that in the corpus as a whole, messages containing more than one language are quite rare - only 3.5%, perhaps a surprising result in a multilingual country but it is in line with our census findings. In the group considered, the low rate of multilingual messages may also be due to a register difference, as there was more code-switching in the meetings. This low percentage of mixed-language messages changed only slightly over the four calendar years, reaching its peak in 2002 (5%). In the data as a whole, mixed messages in which one language is dominant are slightly more common than those in which the two are balanced, but overall the incidence of both is extremely low.

Table 3.3. Distribution of monolingual versus mixed-language messages over time

	% of monolingual messages	% of mixed balanced messages	% of mixed dominant messages	Total Ns
1999	97	0	3	64
2000	97	1	2	235
2001	97	0.3	2.7	332
2002	95	3	2	249
Total	96.5	1.1	2.4	882

Despite the low number of these mixed-language messages, it is worthwhile to examine them more closely, to see what they can tell us about the emerging status of English.

### **"Balanced" versus Mixed-dominant Messages**

A closer analysis of the two mixed types is shown in Tables 3.4 and 3.5. The messages were classed according to how the mixing occurred; whether one of the

---

languages was used for a specific person, whether a single point was translated or whether it was for another reason. Messages classified as "translation" in the tables are those in which the additional language(s) did not provide any new information but rather, merely translated the accompanying text (as in example 5). "Person specific" messages are those in which the change from one language to another appeared to be motivated by a reference to a specific person (as in example 4). "Other" messages are those for which the cause of the code switch could not be determined.

Table 3.4 shows that there were only ten mixed-balanced messages, of which five merely presented the same content translated into one or more languages.

Table 3.4. Distribution of "balanced" bilingual or multilingual messages by languages

"Balanced" messages	French/German	English/French/German	German/English	English/French/German/Italian	Total
Translation	4	1	0	0	5
Person specific	0	0	1	0	1
Other	1	1	0	2	4
Total	5	2	1	2	10

The fact that a number of e-mails provide the same information in more than one language underlines the emphasis members place on ensuring that messages are easily understood by all. Despite an extremely low number of messages of this type it appears that it is when English is not used that this process is important as 4 out of the 5 'translation' messages have only French and German.<sup>38</sup> Of the ten balanced messages in Table 3.4, 5 include English.

Mixed-dominant messages (Tables 3.3 and 3.5) are also very infrequent in the data (N=21). The code switch in many of these messages appears to have been motivated by an intention to address single comments to specific people.

<sup>38</sup> The one 'translation' message which uses English is example 5, and to some extent it could be argued that this was for comical effect rather than for true translation.

Table 3.5. Distribution of mixed dominant messages in two or more languages by dominant language

Dominant language <sup>39</sup>	German+ French	German+ English	French+ English	English+ French	English+ German	English+ Italian	English+ French+ German	Total
Person specific	0	0	0	4	3	2	0	9
Other	3	2	2	4	0	0	1	12
Total	3	2	2	8	3	2	1	21

Although these two groups of messages are different from monolingual ones, the overall trend is the same: English is the language chosen most often. The case of the dominant language messages (Table 3.5) is even clearer; out of 21 messages, 18 include some English and 14 have English as the primary language.

As we have shown that the mixed language messages are not very different from the monolingual ones in terms of language choice, all three groups are considered together subsequently.<sup>40</sup> The mixed messages do, however, show that a certain degree of code-switching is present in communication among association members, and underline their sensitivity to the linguistic needs of a multilingual audience.

Moreover, in terms of the present analysis, note that monolingual messages account for nearly all the data. The next section will consider what proportion of these messages were in English overall opposed to those in French or German, but crucially, as was found for the census, the IFMSA students tend to use a single language, not several, at least within the same message.

### 3.3.7 Change over Time

The two figures below present the results for how English gained importance on the mailing list, one (figure 3.1) presenting the data in terms of the number of messages sent in each language per month, one (figure 3.2) presenting the data in terms of the percentage of language use over six-month intervals. Although the figures essentially present the same data, it is more revealing to examine them together, as in this way

<sup>39</sup> The first language is the dominant one.

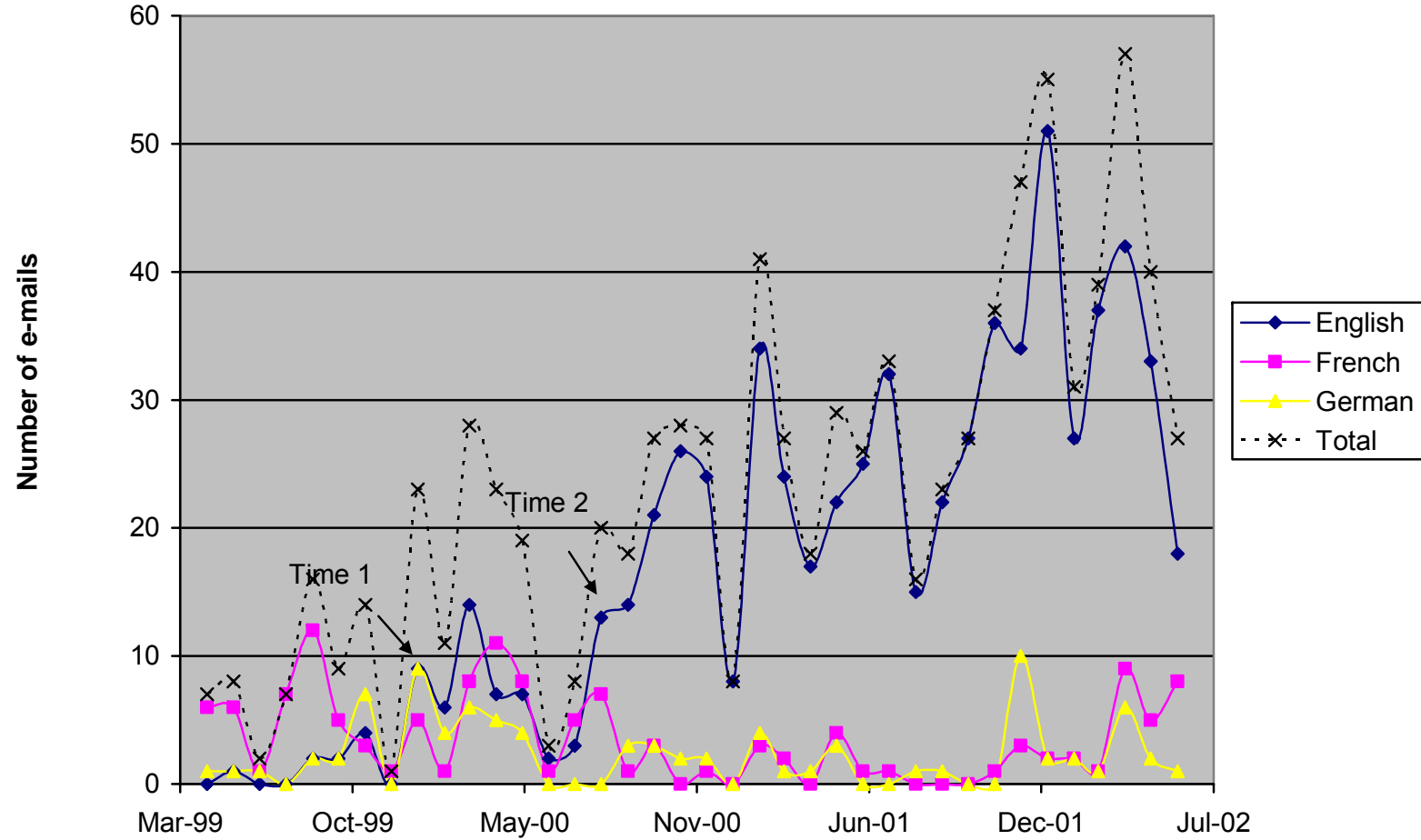
<sup>40</sup> The mixed language messages were coded for the main language used, which was generally possible even in the case of the balanced messages.

they show the increase of messages over time on the mailing list and the crossover to English as the main language but also precisely how prevalent the use of English is.

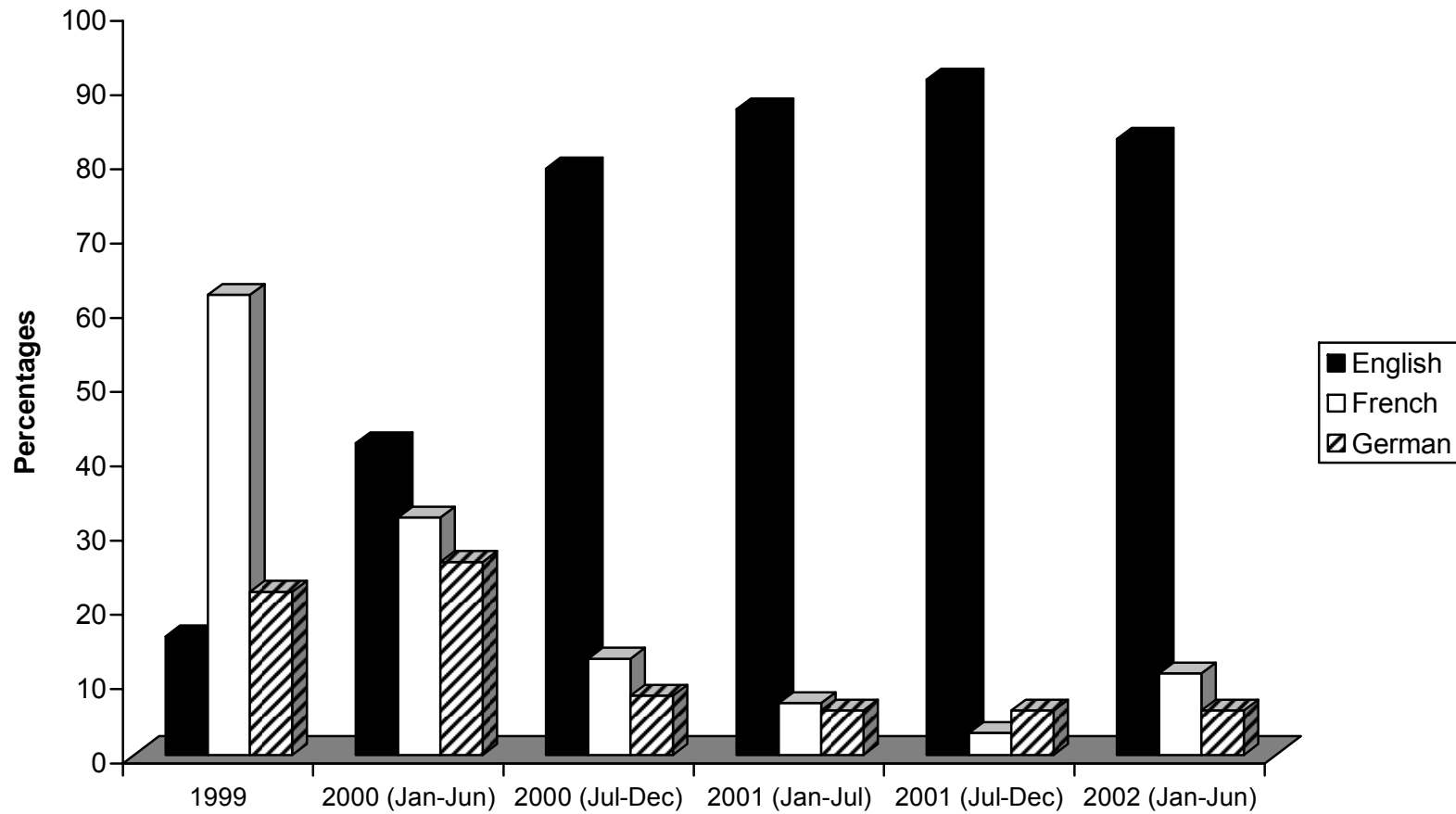
The data conclusively shows that English has superseded the other languages as the main language of e-mail communication for this association. Over a period of about three years, English went from being used a little over 10% of the time to over 80% of the time, with the average percentage in English for the entire period 75% as messages were sent more frequently after the first year (Tables 3.6 and 3.7 below).

Recall from table 3.2 that between 2000 and 2001 there was a shift in the membership of the association (in terms of e-mailers at least); whilst French speakers had sent over half of the e-mails in 1999, the proportion of German and especially Italian speakers increased considerably in the following years. It is in this period specifically that English came to be used more on the mailing list. Undoubtedly, this is due to the reorganization of the e-mailers; the shift to English occurred when French speakers were no longer in the majority and a language that suited the entire association needed to be found.

**Figure 3.1: Spread of English  
(E-mails by month)**



**Figure 3.2: Spread of English  
(E-mails over time)**



Whereas French was the main language in the first e-mails (over 60% in 1999), and German came in second, French and German are used at similar rates in 2002 and English is the main language used. The change in percentages from the first part of 2000 to the second part of the year (signalled by time 2 on figure 3.2) is especially dramatic: English changed from being the main language used to virtually the only language used. The reasons why the change might have occurred at this point will be discussed below. The high use of French in the first set of e-mails is explained by the high proportion of e-mails sent by French speakers in 1999.

As well as the change of language use on the mailing list, the data was also considered in terms of the authors of the e-mails. This was to verify that the change was not due to one speaker, or a group of speakers, using different linguistic strategies than the other members. The data revealed that the change was a global one and that virtually all members used English in their e-mails.

The following Tables (3.6-3.9) present the linguistic choice in terms of the individual speakers; the three linguistic groups are first considered individually (Tables 3.6-3.8) and then compared with each other (Table 3.9).<sup>41</sup> The e-mails in the tables are broken down by calendar year and the number of e-mails sent by the individual members in English is given first, the number of e-mails sent in that period in another language is given in brackets.

Table 3.6: E-mails sent by Italian speakers

Speaker	Gender	University Language	Percentage of English use	No. of emails sent in 1999	No. sent in 2000	No. sent in 2001	No. sent in 2002 (through June)	Total
a	f	German	92%			27 (5)	39 (1)	66 (6)
b	f	French	91%	5 (7)	32 (5)	103 (2)	53 (4)	193 (18)
c	m	German	88%		44 (4)	48 (9)	22 (3)	114 (16)
V	f	German	75%			1	2 (1)	3 (1)
Total	n/a	n/a	90%	5 (7)	76 (9)	179 (16)	116 (9)	376 (41)

<sup>41</sup> E-mailers with fewer than two messages in English were placed in the miscellaneous category

Table 3.7: E-mails sent by German speakers

Speaker	Gender	Percentage of English use	No. of emails sent in 1999	No. sent in 2000	No. sent in 2001	No. sent in 2002 (through June)	Total
d	f	100%			5		5
h	m	73%	(10)	23 (11)	43 (5)	7 (1)	73 (27)
m	f	87%			6	7 (2)	13 (2)
n	m	100%			5	1	6
o	m	88%			7 (1)		7 (1)
p	f	80%			10 (1)	6 (3)	16 (4)
r	f	100%			5		5
@	f	16%	2 (5)	(5)			2 (10)
Total	n/a	73%	2 (15)	23 (16)	76 (7)	21 (6)	122 (44)

Table 3.8: E-mails sent by French speakers

Speakers	Gender	Percentage of English use	No. of emails sent in 1999	No. sent in 2000	No. sent in 2001	No. sent in 2002 (through June)	Total
e	f	44%	(10)	23 (21)	5 (5)		28 (36)
f	m	90%		(1)	11 (3)	35 (1)	46 (5)
g	f	86%			6 (1)		6 (1)
j	m	100%			6	4	10
k	m	100%			5		5
l	m	100%		6	8		14
q	f	100%			2		2
s	f	100%			2		2
u	m	25%	(1)		2	(5)	2 (6)
w	f	78%			1 (1)	6 (1)	7 (2)
&	m	68%	3 (4)	10 (2)			13 (6)
*	f	16%	(15)	6 (16)			6 (31)
R	f	45%				5 (6)	5 (6)
S	m	86%				6 (1)	6 (1)
Total	n/a	62%	3 (30)	45 (40)	48 (9)	56 (14)	152 (93)



Table 3.9: English in e-mails by speaker language

	Percentage of English use	No. of emails sent in 1999	No. sent in 2000	No. sent in 2001	No. sent in 2002 (through June)	Total
Italian	90%	5 (7)	76 (9)	179 (16)	116 (9)	376 (41)
German	73%	2 (15)	23 (16)	76 (7)	21 (6)	122 (44)
French	62%	3 (30)	45 (40)	48 (9)	56 (14)	152 (93)
Misc. <sup>42</sup>	28%	(2)	3 (23)	1 (3)	10 (8)	14 (36)
Total	75%	10 (54)	146 (89)	294 (36)	203 (43)	653 (222)

Although, as the tables show, a majority of the e-mails were written by a small group of people, the rest of the e-mailers make the same linguistic choices. The above tables also show the extent to which individual members change their usage. In the earliest messages most members use their native language (or university language). However, in the later messages, they use English more and more, until they reach a point when they use English exclusively, answering even messages written in French or German in English. Indeed, almost all of the members who wrote in the first and second years start to use English more than French or German by the second year. For example, 'h' wrote 10 messages (out of 10) in German in 1999, but in 2000 wrote 23 in English and 11 in German.

There are two possible reasons for the use of English even when answering an e-mail in French or German. The first is that often, although the original message may be directed mainly to one person, its contents could be useful to all members of the association and should therefore be in a language all can understand. The second reason has to do with the importance of English: these e-mails are in many ways an ideal place for the students to practice their English language skills; general assemblies of the world-wide association are conducted in English, and any communication with members from other countries at these meetings is likely to be in English. As one member wrote,

(6) [i]t is important that you know that you are not "obliged" to write in english if you don't want to, and feel more comfortable in another language.

**But as 'M' said it is a matter of practicity and speed. And you can also**

<sup>42</sup> This comprises speakers of all languages who had fewer than two e-mails in English overall.

**take it as a training to understand the language used in the international meetings and over the various IFMSA mailinglists (if you don't know what I am talking about, visit [www.ifmsa.org](http://www.ifmsa.org))** (b, Italian, e-mail, emphasis in original)

Having determined that English is the most used language for e-mailing in IFMSA, what remains to be seen is whether this situation is held throughout the three language groups, as this will help us determine if the linguistic choices of French, German and Italian speakers are similar. Table 3.10 reveals that English is the main language choice for all three groups. At the same time, Italian speakers write in English more often than the others, probably because they do not have the option of writing in their native language. Moreover, although the Italian speakers belong to a linguistic minority, they are the most frequent e-mailers. However, their frequent e-mailing has more to do with their importance within the association than their minority status ('a', 'b' and 'c' held the three main positions within the association from 2000 to 2002).

Table 3.10 Language of emails by native language of writer

Native language of writer	% in English	% in French	% in German	Total number of e-mails
French	60	37	3	262
German	65	4	31	209
Italian	89	7	4	411
IFMSA Total	75	15	10	882

There are also some differences in gender, as shown in Table 3.11. The results show that female e-mailers are more likely to use their native tongue than male e-mailers; if there is an option to do so. Recall that Italian speakers cannot do this because most German and French speakers do not learn Italian.

Table 3.11. Language of e-mails by gender and native language of writer

Gender of writer	No. of speakers	% in English	% in French	% in German	Total N
Female		71	20	9	503
French	12	44	52	4	149
German	12	51	9	40	86
Italian	4	90	8	2	288
Male		81	7	12	359
French	10	81	17	3	113
German	9	76	0	24	123

Italian	2	87	6	7	123
---------	---	----	---	---	-----

These results are somewhat misleading in terms of the female French speakers, however. The high proportion of French use for this group is due primarily to two speakers who were very active in the first year of data collection, before English had become important on the mailing list so the results need to be interpreted with caution.<sup>43</sup>

What these two tables reveal is that when members choose not to use English, they use their own language. Italian speakers, who do not have the option to use their native language on the mailing list when they chose not to use English, tend to use the language which is spoken at the university they are attending.

### 3.3.8 Causes of Change

What are the reasons for the shift on the mailing list towards a greater use of English over time? One possible explanation is that there was a dominant or core group of e-mailers who led the shift through their linguistic choices. This could be linked to the association acting like a social network or a community of practice (Milroy, 2002, Meyerhoff, 2002); as discussed, Milroy found that dense, multiplex networks work as strong norm enforcement mechanisms; if the main members of the association used English, it may have influenced the others. Tables 3.6 and 3.10 show that Italian speakers contributed the largest number of messages (N=411), and that they were most likely to write their messages in English.

It is not surprising that the Italian speakers use English in such high proportions. More than the others, they would have been sensitive to the need to ensure wider understanding: whether messages were in French or German, they would not have been in the Italian speakers' native language. Those members would have understood first-hand the need to ensure that the messages were readily accessible to everyone. This is not because the Italian speakers could not cope with other languages. Of the three groups, it is the Italian speakers who appear to be the most multilingual. This is unsurprising as they need to be linguistically competent enough to attend medical school in one of the other national languages.

---

<sup>43</sup> Of the 78 e-mails sent in French by female French speakers, 62 were sent by these two members. It is very important to keep these underlying skewing effects in mind when interpreting the results.

The results show that the switch to English was not due solely to the influence of the Italian speakers, however. Not only is English used by all the linguistic groups, but table 3.9 demonstrates that the French and German e-mailers switched to English at the same time (in 2000) as the Italian speakers. Although the Italian speakers sent more e-mails than the other groups, the change to English was widespread. Of all the members who e-mail more than once or twice on the mailing list, only one never uses English (this member is included under "misc." in Table 3.10); every other member uses English at some point. It is also noteworthy that most members who only send a few e-mails use English (e.g., d, j, k). Given these facts, it does not seem likely that it is solely because of the Italian speakers that English has become the main language used; English is the preferred language for all three language groups.

However, it could also be considered that although the Italian speakers might not have overtly imposed English as the main language onto the other e-mailers, their use of English, at a point when the mailing list was starting to be used more frequently, very possibly provided a model for the other e-mailers to follow and as such their role as the first dominant users of the mailing list cannot be discounted.

In terms of cross-linguistic communication, it may be advantageous for the addressees that the writer uses a language that is not his native one as it ensures a greater chance of comprehension. To put it simply, a French speaker will be more likely to understand a German speaker's basic English than the same speaker's native German (if only because the German speaker lacks the knowledge to make use of more complicated forms in English). In the interviews, as well as in a number of the e-mails, the members commented that they were aware that their English might not be as rich as a native speaker's but that it was still suitable for communication.

(7) I think er that the English er which, no – which is used all around the world just to communicate er for the er basis of communication it's a very poor English, I mean, I really don't have a lot of vocabulary in English [...] it's incredible, that's unbelievable in English is that **we can speak a lot but we use always the same words.** (w, French, interview, emphasis mine)

Another reason why it is likely to be beneficial to use a non-native language is that this will put all the speakers at an equal disadvantage; if the French-speakers and the German-speakers used both French for example, then the French-speakers would be more at ease of course.

The general feeling in Switzerland about the position of English vis-à-vis the Swiss national languages must have played a part in the language choice as well. IFMSA-CH's choice to use English as the main language of intranational communication is in no way unusual for the country. English is seen by many in Switzerland as a necessary tool to further oneself in society, as the decision of the canton of Zurich to make English, and not French, the second language taught in schools demonstrates (Busslinger, 2004, Cossy, 2005). The language choices of the IFMSA-CH students are consistent with this trend.

English is also the language of the medical sciences world-wide: many of the important journals and textbooks are written in English, and even within Switzerland many medical research groups use English. In the European Union (of which Switzerland is not a member), it has even been said that:

[t]he trend of using English as the teaching medium at continental universities can be seen in almost all scientific areas. It often applies only to single courses, and occasionally to a whole degree....The trend towards an entire degree being taught in English is most visible in business schools.  
(Phillipson, 2003:77)

So the members would have been aware of the importance of knowing English for their studies. Note however that medical instruction in Switzerland is not in English. Some students can, and indeed do, manage without speaking English at all. Medical students in Switzerland attend medical school as undergraduates, whereas it is primarily in postgraduate studies that English is used in Swiss universities (Dürmüller, 2001).

All of the above reasons played some part in the language selection on the mailing list. However, the reasons the members themselves give as to why they chose to use English suggest that in many ways the choice ultimately had more to do with language comprehension than with anything else.

### 3.3.9 Language Comprehension

Evidence for the importance of English is found not only in the number of e-mails written in English, but also in what is said in those e-mails. The e-mails provide a plausible explanation at the very point when English begins to be used more frequently than the other available languages (time 1 in figure 3.1). In 2000, when the members needed to decide who was going to the General Assembly, and the president was away in New York and could not give much advice, a misunderstanding occurred because some members had already registered, but found they might have to give up their place so that someone from another university could attend. The president sent the following e-mail in French:

(8) En gros, si Basel a déjà fait le transfer on a 9 personnes plutot que 8, mais si ce n'est pas encore le cas alors on perd notre place en plus... Donc si on a 8 place, je pense que soit 'h' soit 'X ' et soit 'u' ou 'e' ne peuvent pas venir... et que Lausanne devrait venir, mais je vous laisse decider... {...} Desolee mais pour l'instant je suis a New York donc ne peut pas beaucoup vous aider... (\*, French, e-mail).<sup>44</sup>

E-mails went back and forth among members in French and German in an effort to determine who was in fact registered, and who should give up his or her place. Finally, one member wrote: "it's about time that things get clear in this meeting story" (&, French, e-mail) and went on to summarize the events as he saw them, in English. The next few e-mails on this subject were in French and German. Then this member wrote again in English, summarizing the discussion and ending his message by saying "Thank you for your answer, so the actual position is:" This helped him clarify the position and make sure he had understood the German, but it benefited other people as well, and other members then thanked him: "vielen dank für die aufklärung!!" (h, German, e-mail —Thank you for the explanation).

This indicates that the messages in French and German were not immediately understandable by both groups. When the president got back, she then began writing in English as well: "I'll continue in English as '&' started... Sorry for all those strange

<sup>44</sup> Translation: Basically, if Basel has already transferred the money we have nine people rather than eight, but if it isn't the case we lose our extra place... So if we have eight places, I think that either 'h' or 'X' and either 'u' or 'e' can't come ... and that Lausanne should come, but I'll let you decide ... {...} Sorry but for the moment I'm in New York so can't help you much...

messages, it would have been simpler if I was in Geneva and I could have phoned you but... life is life!” (\*, French, e-mail). She then summarized the entire discussion once again. After this episode, the member who had started using English continued to use it for most of his e-mails, and many of the others who had previously only used English occasionally started using it more often as well.

Since the main change to English occurred at this time, it is highly plausible that it was in part influenced by these developments. After using both French and German, the IFMSA-CH members had to come to terms with the fact that it was difficult for the French speakers to understand all the German messages and vice versa. English proved to be a more readily accessible foreign language for both groups. One member wrote, "well i' keep going with english, since my french is even worse..." (h, German, e-mail). Within the mailing list, English was the most accessible language choice; speakers were able both to write it and understand it if someone else used it to a far greater extent than other non-native languages.

Another e-mail is also very telling when considering the reasons why English is used so frequently on the mailing list. This e-mail followed soon after the initial switch to English and was sent in response to another member sending out an invitation to a national meeting in French. The subject of the message was "Switzerland is a multilanguage-speaking country." In this message the member admonishes other e-mailers to use English, saying,

(9) please, next time you send a message to IFMSA Switzerland you have to write it **in english**, or to send it **both in french and german**. This thing is even more important when you're sending an official invitation for a national meeting like this one. I think you'll understand why.

(c, Italian, e-mail, emphasis mine)

The message sender was one of the very frequent e-mailers, who consistently followed the practice he advocated, so this undoubtedly contributed to the increase in English. In fact, this e-mail was posted at time 2 in Figure 3.1 after which point French and German e-mails were far less frequent. This e-mail also raises another point: if English is not used, then both French and German are necessary. The feeling that French or German cannot be used alone is also found in e-mails by a member who chooses not to use English, if he can avoid doing so.

(10) Es tut mir leid aber ich kann nicht auf Englisch mit Schweizer kommunizieren... Je suis désolé, mais je ne peux correspondre avec des Suisses en Anglais...(u, French, e-mail - I'm sorry but I can't communicate with/ write to Swiss people in English).

Although this member says that he does not wish to use English, he recognizes that he cannot use only one of the Swiss languages. To ensure total comprehension he must write in both. He does in fact try to send all his messages in both French and German, but he does not e-mail very frequently, possibly because of the effort involved in producing dual-language messages. He does hold one of the top posts within the association, at one point, however.

Similarly, speaker 'c' adds as a postscript to a message which he wrote in German: "Si quelqu'un n'a pas compris mon message, je serai heureux de l'envoyer aussi en français." (c, Italian, e-mail— if someone hasn't understood my message, I'd be happy to send it in French as well). The alternative to English is two languages - democratic, but less efficient.

These situations provide us with tangible clues why English came to be the lingua franca on the IFMSA mailing-list. Moreover, the fact that they occur at points which figure 3.1 had already revealed to be turning points in the change to English underline how relevant these situations were to the linguistic choices of the association.

To summarize the main findings of the linguistic analysis of the e-mails; members initially used their mother tongue (or university language in the case of the Italian speakers) on the mailing list, but they quickly changed to English, making it the *de facto* language of the mailing list. This change was a global one and affected all three linguistic groups. Explanations why this change occurred can be found in the metalinguistic commentary of the e-mailers themselves; not only did English allow the students to practice their English and ensure that messages were understood by all, but it also put the members on an equal footing linguistically as it was a non-native language for all of them.

The association was also found to have many points in common with social networks and communities of practice. This is important as 'close daily interaction in the community' was one of the four agencies which Le Page and Tabouret-Keller (1984:187) discussed as



promoting focussing. What of the other agencies, to what extent can they be applied to the mailing list situation as well?

Of the remaining three agencies, two - ‘a powerful model’ and ‘the mechanisms of an education system’ (Le Page and Tabouret-Keller, 1984:187) – can be related to the Swiss situation, while the third ‘an external threat or any other danger which leads to a sense of common cause’ (Le Page and Tabouret-Keller, 1984:187) does not. The members of IFMSA-CH do not use English because of an external threat, but there is no doubt that the first speakers who used English provided a powerful model for the subsequent e-mailers. Moreover, although the mailing list is not exactly ‘an education system’, it teaches the members which language it is appropriate to use on the mailing list. Members know what language (and what type of language) to use in their e-mails because they have the model provided by earlier messages. If focussing is occurring in Switzerland, we would expect to find signs of it in a group such as this one.

### **3.3.10 Type of English used**

Having established that the IFMSA-CH mailing list can provide us with English data from the three linguistic groups of Switzerland and shown the reasons why English became the main language of the association, this thesis will now be able to examine the English used by the members. This analysis of the members’ English will be dealt with in the second section of this thesis. There are, however, a few more issues which need to be presented at this point in terms of the English used by the IFMSA members.

There is no way to control for the amount of data for an individual on a mailing list; some people e-mail very frequently, some rarely. However, the tables presenting the number of e-mails sent by the speakers in the different linguistic groups (3.6-3.9) show that most members sent at least five messages. This will allow us to obtain an accurate picture of their use of English. Of course, because we have over a hundred e-mails for some students and less than 20 for others, it will be necessary in the feature analysis to always code separately for individual to establish that the members of a linguistic group are similar.

In all the e-mails analyzed in this study, the main concern of the students is to get their point across, not to write perfect English. There is a communicative urgency, rather than a

desire to create flawless sentences, and members are quite willing to admit that sometimes their English is imperfect. When they do not know the exact word in English, they will often replace it with the corresponding one from their own language. Idiomatic forms are translated literally, sometimes accompanied by the phrase, "as we say in French/German."

The students are aware that their English is different from the Standard English they hear on television or the radio or in movies, but this does not trouble them as it serves its purpose. In fact, when the association was first approached to get permission to use the mailing list for study, quite a few of the members were keen to see what their own special brand of English was like. The initial contact with the association also spawned a number of comments by the students on their use of English. For example, when one person commented that he thought it was too bad that they could not use one of their own national languages on the mailing list, another commented that the English that they used, which incorporated structures from the various Swiss languages, might in fact be a type of Swiss language itself:

(11) Je comprends le point de vue de 'c'. Man \*muss\* aber im fall nicht alles auf english schreiben, but it just so happens that this is what we are (naturally?) inclined to do in order to get the widest audience and understanding at once (justement pour ne pas avoir à faire ce que tu as gentilement proposé de faire à la fin de ton message 'c': traduire pour ceux n'ayant rien capté!). **The study might show that we have our own special swissicized variety of english, at which point you can wonder if it isn't also a type of national language, which combines structures, expressions and mindsets of all the other national languages, une sorte de mélange.** Whatever. Bin auch nicht gegen das projekt und finde es interessant. (M, French/English, e-mail, emphasis mine).<sup>45</sup>

### 3.3.11 Final words on IFMSA-CH

---

<sup>45</sup> Translation of the German and French segments: I understand c's point of view. But we don't \*have\* to write everything in English but {text in English} (precisely to avoid having to do what you kindly offered to do at the end of your message 'c': translate for those who didn't get it!) {text in English}, a kind of mix. Whatever. I'm not against the project and I think it's interesting.

While results from a single mailing list cannot be generalized to the linguistic situation in all of Switzerland, they provide suggestive clues as to why English has gained in importance as a lingua franca in Switzerland over the past few years. The main English use of this group is primarily with other Swiss nationals, not with native speakers of English. The reasons for this are quite evident; the members use English primarily on the mailing list, in a situation when they are dealing with members from the whole of Switzerland, the other situation when they use English is at national meetings (if one considers that the one in Valais was not usual).<sup>46</sup>

English appears to be the most readily understood and accepted language in mixed language groups, the main reason for this being that it is a non-native language for all. The Italian speakers on the mailing list were among the forerunners in this change since they experienced first hand the need to ensure that people be able to understand one another, as nobody else spoke their native language. They were not the only members involved in this change, however; as the analysis revealed that other speakers also aimed to ensure that the messages could be understood by everyone from the beginning.

It is likely that the mode of communication, the mailing list, also influenced the choice of language used. As long as communication (face-to-face or otherwise) was mainly directed to specific people, there is no need to worry about widespread understanding. However, when the aim is to communicate to a broader, multilingual audience, as the Internet makes it easy to do, neither French nor German is able to serve as the main language in the Swiss context, and it becomes necessary to use English. Although two or more languages can be used for face-to-face interaction (as was demonstrated at the second meeting), in the case of e-mail, the use of more than one language was impractical and confusing, and precipitated the eventual choice of a single language, namely English.

Undoubtedly, other groups in Switzerland have faced and will continue to face similar problems, as will multilingual groups collaborating online in other cultural contexts. They, too, may choose English as their lingua franca, facilitating cross-language communication and contributing to the increased use of English on a global scale.

---

<sup>46</sup> Comments from members before and after this meeting would seem to indicate this. Both the president and one of the people interviewed informed me that meetings were generally conducted primarily in English.

One difference between the students' choice to use English and that of Swiss companies which use English as their official language is that the members of the association chose it for primarily practical reasons rather than economic ones. The students' use of English is presumably purely determined by the fact that it is the most accessible language for all. Moreover, English is the language of medical science around the world, and should the students decide to pursue medical research then English would be a necessary language for them. Although companies might not choose to use English solely because it is a practical choice, this fact will be a contributing factor in their decision as well.

From the comments of the students on the mailing list as well as those who were interviewed, it does seem indeed the case that the members did not use English outside the confines of IFMSA very often; the mailing list was a place where members could practice English. Even if their English use was somewhat restricted, in that it was primarily used for e-mailing but not often for conversations, the members were quite likely to be in contact with English on a regular basis through the mailing list. There was roughly an e-mail a day sent out to the members. Even if the members did not e-mail themselves, they read messages in English every day and thus would have had a very different feeling for it than for a language (French or German depending on the region they are from) which they only used occasionally. If one considers that it is rather difficult to use a language infrequently especially if one has only learnt it at school, then English should be considered quite differently from other non-native languages.

### **3.4 Native English data**

Recall that this thesis aims to compare non-native speakers' variation patterns with those of native speakers. The native data considered must then be directly comparable to that of the non-native speakers, as differences of register or style between the two groups could easily obfuscate shared patterns. Because the main part of the Swiss medical student data was composed of e-mails, a similar corpus of e-mails sent by native English-speaking students was collected. This native e-mail corpus will be the main point of comparison for the non-native speaker results but, for the features where it is possible, previous results will also be examined.

To construct this native corpus, I obtained permission to use the mailing list of a sports society at the University of York in the United Kingdom.<sup>47</sup> Although the e-mails of the sister association of IFMSA in the UK could not be obtained,<sup>48</sup> the native mailing list selected contained e-mails of a very similar nature to the IFMSA-CH ones. Because the university association chosen was one that I had been directly involved in, I knew that, as in the case of the IFMSA-CH mailing list, e-mails were sent to organize events and provide information for the club's members. The style of the e-mails was, as on the IFMSA-CH list, informal as members knew each other well (see example 12).

(12) Hiya everybody,

We're going to The Garden of India on Wednesday the 27th of February which is this wednesday, and following it up with a visit to the gallery for some boogying!!

I'm going to book it for 8:30 pm so there's to be no faffing about after the session, you'll probably only have time to drop bags off (The Gallery cloakroom wont take bags), so best get changed straight after the session, as we need to get there on time!

The following people are on my list: [List removed]

If there is anybody else who wants to come along, please let me know by 12pm tomorrow at the latest (sorry about the short notice - it's was a bit of a last minute decision).

While the corpus itself is relatively smaller than the Swiss corpus as a whole (around 40 000 words in total) and the number of e-mailers is more restricted as well, the native corpus is made up of one language group unlike the Swiss corpus which contains data from three different groups of speakers.

The purpose of this corpus is to serve as a control group to the IFMSA data, allowing me to compare the results for the non-native e-mail data with native e-mail data. Although they can be relatively informal, e-mails are not directly comparable to speech, nor are they equivalent to writing. Any results from previous studies looking at native speech cannot be compared to the non-native results without first establishing that native speakers do not

---

<sup>47</sup> The e-mails were sent by both members of the association in York and members throughout the United Kingdom.

<sup>48</sup> A quick check online did not reveal a comparable mailing list in the UK.

change their distribution of the forms in another register. For the features where native studies already exist, the native e-mail corpora will act as a bridge between the native oral and non-native e-mail data, verifying that any differences encountered are due to genuine differences between the native and non-native speech patterns and not to register differences.

Computer mediated communication has been found to be somewhat distinct from other types of linguistic data and its place on the continuum of speech and writing has not yet been clarified (Herring, 2001). This thesis will need to be very careful not to equate e-mail data with speech or writing, but rather establish how the patterns of each feature considered are affected by the unconventional register.

### 3.5 Extra data

As well as being a valuable source of data in terms of the mailing list archive, the Internet offered another possibility for the study of English as it is used in Switzerland. To be precise, search engines such as Google allow us to determine how prevalent a feature was in Switzerland. It is quite easy to search for two variants of one feature and restrict one's search solely to web-pages in Switzerland and in English. As will be detailed further on, the plural variant of *information* appeared only in one of the corpora and yet it did seem to be a fairly generalized term. A Google search revealed that the plural form occurred quite often on Swiss web pages in English, nearly as much as the singular and native one. Rather than having to discard a feature because it was only found in one of the corpora, Google allows us to establish how widespread it is.

One of the main difficulties of web-pages, as opposed to mailing lists, is the fact that it is for the most part impossible to know who the writer is, and accordingly what language might be the mother tongue. This is why the search engines can only be used to complement already existing data and not as a source of features on its own. Even so, generally enough tokens can be found to determine that the feature does not belong solely to one group but possibly to other English speakers in Switzerland.

Internet searches will also enable us to verify whether the features might also exist in other non-native varieties of English, by circumscribing a search to a defined country we will be able to see if a feature is present elsewhere as well. This should not be taken as a substitute for an actual analysis of the feature in other linguistic situations but at times when that should prove impossible or too difficult it can at least serve as a direction for further research. For example, a Google search on French websites in English looking for the term ‘informations’ provides us with about 50’000 hits whilst the standard term ‘information’ comes up with 100’000. From this we know that it is not only the Swiss speakers who use the plural with non-count nouns. Of course, this method does not allow us to determine whether the French speakers use the same percentages and patterns as the Swiss speakers. It also permits us to determine to what extent a figure of speech might be idiosyncratic or language restricted; one of the Italian speakers occasionally uses ‘I please you to do something’ – an expression most likely calqued from ‘vi prego di’ in Italian – by searching on Google it was determined that she is not the only speaker to use it, but also that it is found in other Italian speakers as well as speakers from other languages (Slovenian for example).<sup>49</sup>

### 3.6 Discussion

The IFMSA-CH association has been shown to be an ideal ground to study focussing and non-native variation; not only has it been possible to establish why the medical students have chosen to use English rather than an indigenous language, but the association has the markings of a close-knit social network. This type of network is precisely the kind associated with focussing (Le Page and Tabouret-Keller, 1985:187). The fact that the non-native speakers hail from three different linguistic backgrounds (French, German and Italian) will allow us to consider the affect of the source language in the acquisition of systematic variability in English. By examining data from a mailing list, we will be able to lessen the effect of the Observer’s Paradox (Labov, 1972) and ensure that we are considering English data as it is genuinely used by the IFMSA-CH members.

---

<sup>49</sup> A Google search 08/09/2006 of the terms “I please you to” yielded 37900 hits.  
(<http://www.google.co.uk/search?q=%22I+please+you+to%22&start=0&ie=utf-8&oe=utf-8&client=firefox-a&rls=org.mozilla:en-US:official> )

The native data, which was selected as a control group, has a number of points in common with the non-native data. First of all, the native corpus is composed of e-mails, so the two groups share the same register. This will allow us to compare the patterns in two groups. Secondly, the e-mails sent by the native association are of a similar type as those on the Swiss mailing list; in both associations members know each other fairly well and the style of the e-mails is informal.



## **4. Methodology and variable selection.**

### **4.1 Introduction**

This chapter presents the features and the methodology that will be used in the subsequent analyses. One of the selected features will be used in this chapter to introduce the methods used for analysis in order to make them as understandable as possible for the reader. This will provide a more practical demonstration of the methodology introduced in section 2.5. To accomplish this, the chapter has been divided into two main sections; the first introducing the features which will be dealt with in the present thesis, and the second section focussing on the processes used to analyze the various features. I will also outline how the outcomes of the various analyses will be used to establish whether:

- the non-native speakers have acquired native variability
- the non-native varieties have focussed into a single variety.

A feature involved in focussing will have very different results from one which demonstrates that native-like variability in the three non-native groups (figure 4.2). We will thus be able to use these results to test which of our initial hypotheses was accurate.

### **4.2 Feature selection and presentation**

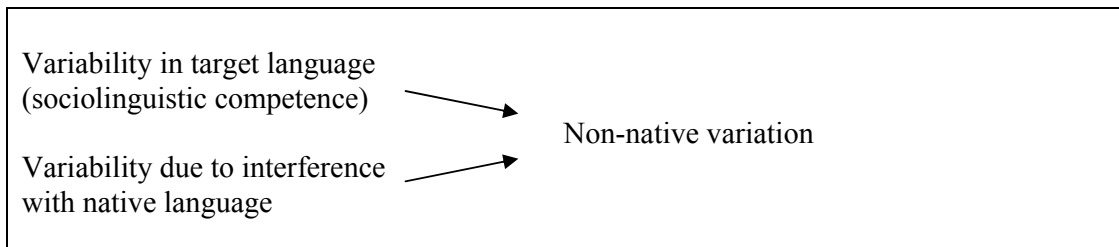
#### **4.2.1 Feature selection**

To determine whether the English of Swiss speakers with different mother tongues is similar, we will need to examine a range of features. A single feature can give us only limited insight into the workings of a language, so by considering several features we will be able to determine far more about the processes of non-native variation. Different features yield different results and conclusions on the state of English in Switzerland and it is their combined analysis which will allow us to establish what processes are under way. A range of features will also help us establish how non-native speakers deal native variation in general.

While there is a primary type of non-native variability which is largely linked to the source language and to linguistic interference (see examples 1-3 and detailed discussion below), non-native variability types which are linked to the target language and are linguistically conditioned must be considered as well (see examples 4-5) (Bayley and Regan, 2004:325).

As mentioned in chapter 2, one reason why second language acquisition researchers have tended to dismiss variability in non-native varieties is that they consider only one type of non-native variation; cases where non-native production is at odds with native speaker production. However, as demonstrated by a number of studies into non-native speakers' sociolinguistic competence (for example, Bayley and Preston, 1995, Rehner, Mougeon and Nadasdi, 2001, Mougeon, Rehner and Nadasdi, 2004), there is a second type of variation which is noticeable only when it is studied both quantitatively and qualitatively; cases where non-native speakers have acquired the variability already existent in the target language.

The difference between these two types of variation needs to be made explicit, as it will help justify the selection of features in this thesis. Although some research on second language variation has in fact examined the variability associated with the target language, there has been little attempt to categorize the two different types explicitly.<sup>50</sup> This thesis considers non-native variation to come from two sources (see Figure 4.1).



**Figure 4.1: The two sources of non-native variation**

### **Interference-based variation**

I have designated variation linked to interference from the source language or with general problems of acquisition as Interference-Based Variation (IBV). In cases of variation of this type, non-native speakers sometimes produce a construction in a grammatically inaccurate way (example 1) and sometimes as it would be produced by a native speaker (examples 2-3).

<sup>50</sup> There has been some categorization, however. Regan (1995:178), for example, discusses variation in interlanguage as being on two axes, 'vertical and horizontal.' Considered in this manner 'progress along the vertical axis can be seen as progress along the developmental line in the acquisition of native structures of the language whereas progress along the horizontal axis is seen as the acquisition of sociolinguistic competence or the acquisition of dialect differences' (Regan, 1995:178).

- (1) In Geneva, we command our new computer, with a scanner **who** is also a fax and a photocopieuse (colours and black and white)!
- (2) For this, I write here the way for helping students **who** want to stop to smoke.
- (3) when you find an activity **which** could replace smoking (all f, French, e-mail)

In the examples above, the variation in terms of relative pronoun selection is linked to what is permissible in French; relative pronoun choice is not governed by whether the subject is animate or not in French, whereas it is in English. *Who* is an acceptable pronoun for *students*, but not for *scanner* and *which* can be used only for inanimate subjects (such as *activity* or *scanner*), while in French, however, *qui* would be the relative pronoun used for all three. The fact that the distinction does not exist in French most likely led the French native speaker to sometimes use the form reserved for animate subjects with inanimate ones (as in example 1). This represents an example of variation rather than simply being an error or a mislearned rule, however, as there are also cases, as in examples 2 and 3, where the Standard English structure is used instead. Cases of interference-based variation are, on the whole, not very difficult for researchers to discern, as one of the variants will be grammatical to native English speakers, while the other one is not.

### Target-based variation

For the purposes of the present study, I have designated variability which derives from the target language as Target-Based Variation (TBV), as in the examples below:

- (4) The first people **that** come up with a german and a french version, please send it out
- (5) I am sending it in advance, so that you can prepare yourselves and also the people **who** will not be in Bern are able to express their thoughts! (both b, Italian, e-mail)

In the above examples, the same subject (*people*) of a relative clause is used with two different relative pronouns by the non-native speaker. This is due to the fact that Standard English allows for the use of both *that* and *who* with the subject *people*. Non-native speakers, such as the ones in IFMSA-CH, will have been exposed to both variants while learning English and thus have both forms in their repertory in their own English use. In

this instance, the researcher is faced with two variants, both of which could be used by a native speaker of English.<sup>51</sup>

Given the differences between the two types of variability, the interpretation of their results and their analysis will be different as well. In order to determine whether focussing is underway in an interference-based variable feature, I will compare the French, German and Italian results and assess the extent to which the non-native tokens in each separate language group are influenced by the same factors. For target-based variables, I will look, first of all, at whether non-native speakers are following the same patterns as native speakers. If it is determined that they are not, then an examination of whether the feature demonstrates focussing or individual variability will be carried out. Multivariate analysis can be used with the two types of variable, but native variability will only be considered in cases where TBV is also found.

Once the more quantitative aspects of the analysis for each of the four language varieties (English and the three non-native languages) are completed, the feature as a whole will be considered using the methodology of comparative sociolinguistics. This will enable us to establish if a feature has been acquired natively, if it demonstrates signs of possible focussing, or if it is neither of these two options as schematized in figure 4.2 below.

Case	Patterns Found	Outcome	Types of variation it is found with
Case A	French, German and Italian present similar variation patterns to native speakers	variation acquired	TBV
Case B	French, German and Italian variation patterns are unlike native speakers but similar to each other	focussed form (or common problems of acquisition)	TBV, IBV
Case C	French, German and Italian vary unlike native speakers and unlike each other	variation influenced by native language only	TBV, IBV

Figure 4.2 Potential cases of variability

<sup>51</sup> This feature will be discussed further in chapter 6.

The three variation pattern types for individual features will be determined as follows:

- CASE A

A feature will be categorized as case A (native-like acquisition of variability) if an analysis, using the methodology of comparative sociolinguistics, can conclude that the patterns of variation in the English of all three language groups are closely comparable to each other and to the native speaker control group. In such a case, the assumption will be that the non-native speakers have the same variable rules as native speakers.

- CASE B

Variables that do not pattern like native speaker variables, but demonstrate that all three language groups vary in the same way are categorized as case B. This category will demonstrate the presence of a potential Swiss English variable. If a variable patterns in the same way for all three language groups then it is a likely point of convergence, especially if it can be demonstrated that the patterning does not match what would be found for non-native speakers of the Swiss languages in countries (i.e. France, Italy and Germany) where English does not have a comparable lingua franca use.

- CASE C

Variables in which none of three language groups have the same patterns in English and furthermore none of them present the same patterns as native English are categorized as case C. In this case there has been neither acquisition nor focussing but simply linguistic interference of the source languages. In cases such as this one, the variation may be turn out to be individual or free and speakers from the same linguistic background may not have the same patterns.

Features belonging to case B are considered to be potential rather than bona fide features of a single variety of Swiss English, because the patterns found might be shared with other non-native varieties of English as well. The identical patterns could be indicative of other processes, such as pidginization, general difficulties in learning a specific aspect of English syntax (common to all non-native speakers and not just the Swiss speakers) or due to some

type of linguistic universal at work. Features found in case B are undoubtedly cases of focussing but determining whether the focussing is restricted to Swiss speakers is beyond the scope of this thesis.

Target-based variation features can belong to any of the three categories, while interference-based variation features will only be found in case B and case C.

#### 4.2.2 Presentation of features

Variables have been selected on the basis of their type of variability (interference-based or target-based), frequency, and whether (for target-based features) they had previously been analyzed in native varieties of English. This last was important as it will allow us to compare the results of this analysis with previous studies.

The rest of this section provides a brief overview of the features which were chosen as suitable for further study and which will be dealt with in the subsequent chapters. Because this study focused on a relatively new area of SLA, i.e. the acquisition of inherent variation, the bulk of the analyses focus on features which are primarily linked to target-based variability.

##### A. Interference-based variation features

###### 1. *information* vs. *informations*

The section below which deals with this feature examines the use of *information* as a count noun rather than as a non-count noun in the English spoken in Switzerland (examples 6-7), and aims to uncover whether the factors influencing the standard or non-native use of the form are similar across the three language groups.

(6) Please, send to me the missing *informations* requested in my report

(7) I am sorry so sorry.... I didn't know this *information* (both b, Italian, e-mail)

It highlights the importance of studying native language interference quite closely to ensure that the apparent similarity between groups is not due merely to equivalent structures in the three source languages. In this case, the non-count noun *information* is a count noun in the three Swiss languages. This feature is not a variable of the kind that is generally studied by

sociolinguists, but rather the kind considered by second language acquisition researchers. The analysis will demonstrate how the aims of sociolinguistics and SLA can be combined even in cases where one of the variants is non-native.

## B. Target-based variation features

### 2. *will* vs. *going to*

The chapter on the future examines one of the aspects of variation in Swiss English which permits comparison with native English, namely the variable use of the future forms *will* and *going to* (examples 8-9). Although there are differences in the distribution of *will* and *going to* in different native varieties, including a difference of overall distribution of the two variants between British and North American English, many of the factors which condition the variability have been found to be shared across varieties (Tagliamonte, 2002:751ff). If the Swiss speakers are following the native patterns then an analysis of the factor groups involved in the distribution will demonstrate this.

(8) I *will* try to keep the facts up to date (o, German, e-mail)

(9) this *is gonna* be a sweet ride on the night train (o, German, e-mail)

Unlike some of the other target-based variation features, both of the variants for the future were explicitly taught. This will allow us to establish whether overt teaching of variability has an effect on non-native patterns.

### 3. relative pronouns

This chapter deals with the selection of relative pronouns by non-native speakers. It considers the choice between the pronouns, *who(m)*, *that*, *which* and the zero form (examples 10-12).

(10) but send me a list of people *who* want to go there (\*, French, e-mail)

(11) or at least the people *that* might concerne it! (k, German, e-mail)

(12) and the number of people  $\emptyset$  you are paying for. (h, German, e-mail).

Although there are a number of prescriptive rules which affect the use of the various relative pronouns and these are taught in Swiss schools, many of the underlying effects of variation (some aspects of animacy and syntax as well as adjacency and definiteness) are not taught. This analysis will allow us to see if these underlying factors are nevertheless affecting non-native speaker patterns of relative pronoun selection.

#### 4. complementizer *that*

The selection of complementizer forms also provides speakers with the option of either using an overt form (*that*) or a zero form (examples 13-14). While a preliminary analysis revealed that the non-native speaker groups use both variants, the analysis aims to establish whether this feature's patterns are similar for the native and non-native English speaker groups.

(13) I think  $\emptyset$  we will be 20 medical students (f, French, e-mail)

(14) I think *that* he will be great in helping you (b, Italian, e-mail)

Native speaker use is affected by factors such as lexical verb, grammatical person and tense (Tagliamonte and Smith, 2005: 301). Non-native speakers are not taught that these factors affect variability; in fact the two variants are presented largely as free variation. This chapter will consider whether the fact that there is no overt teaching of the variability will hinder non-native speakers' use of the two variants.

These three target-based variables will provide us with valuable data on their own but considered together they will also allow us to consider a further point; whether overt teaching of variability facilitates the acquisition of the native-like patterns. While the variation in *that* and zero complementizers is not taught at all, the variation in relative pronoun use is only partially taught and finally the variation between *will* and *going to* is presented to some extent. Will we be able to see an affect of this teaching in the results?

#### 5. Additive adverbials *also*, *as well* and *too*

This chapter will focus on lexical variation that is found in native speakers, namely the variation between the three additive adverbials *also*, *as well* and *too* (example 15-17).

(15) I think the grammar is changing **as well** now (S, French, interview)

(16) Actually, this message is **also** a big HELP request (j, French, e-mail)

(17) it was a situation that I shared, **too**. (f, French, e-mail)

The variation of the three forms has been found to be linked to style and to register. Furthermore, one of the variants, *also*, is acceptable in a wider range of positions than the other two forms which will allow us to uncover whether non-native speakers were likely to extend the use of the other two variants to positions not acceptable to native speakers. The analysis will consider both whether the non-native groups share the patterns of native speakers and whether the three groups have similar problems in terms of placement.



This chapter will help establish whether there are differences between target-based and interference-based lexical variation, by offering a counterpoint to the section examining *information* as a count noun.

### 4.3 Lexical Variation I: Informations

Having presented the features which will be analyzed, this section will, taking the first feature, provide a sample analysis which will guide the reader through the various methodological steps. Moreover, by taking the one interference-based variation feature selected for analysis, this section will also be able to reveal some of the inherent flaws in looking solely at variation due to non-native interference in a study with aims such as this one.

Each analysis has been divided into four main sections. The first section of the feature analysis focuses on a **description** of the variable selected. This section deals with how the feature is used in Standard English, consulting descriptive grammars of English (such as Quirk, Greenbaum, Leech and Svartvik, 1985) as well as previous research on the feature. It also deals with the variants which are found in the non-native speakers. This section will also present the source language counterparts (if there are any) for the feature. This will involve an examination of the equivalent syntactic constructions in the Swiss languages to determine what factors may account any differences between language groups.

The second section will focus on **the extraction and coding** of the tokens. As well as presenting any previous studies conducted on the feature, this section will detail the factors which have been found to account for aspects of the variation (both in Standard English and in non-native English). It will also serve to report on any tokens of the feature which were excluded from consideration.

The third section will concern itself with the **analysis** of the feature. Whenever possible the analysis will be divided into two main parts. The first part will present the comparison of the marginals, that is to say the ‘frequencies and percentages of the variant forms in the data according to independent variable(s)’ (Tagliamonte, 2006:265) and the second part will discuss the multivariate analysis results. There were a number of cases when it was not

feasible to conduct a multivariate analysis, either due to a low number of tokens or due to there being insufficient variation (Tagliamonte, 2006:182, Guy, 1980:30, see individual variables for a further discussion of this). This section will present the results for native speakers and non-native speakers. The analysis will focus on aspects of target-based variability but also on interference-based variability. The comparison of the marginals will attempt to determine whether the feature has been acquired in the same way for the non-native speakers as for the native speakers. The multivariate analysis results will present the factor groups found to be significant and establish if the hierarchies between the different speaker groups are similar. Using the methodology of comparative sociolinguistics, the comparison of the speaker groups will allow us to establish if the feature presents native-like distribution, focussing or free variation. This will be discussed in the final section of each analysis, which will focus on **categorizing** the feature. This final section will also offer a conclusion.

#### 4.3.1 Introduction

A preliminary data analysis of potential features of a focussed Swiss English revealed a number of cases of problematic plurals and, more specifically, cases where non-count nouns were used as count nouns. Examples 18-20 show that the noun *feedback* is used as a count noun by the Swiss speakers, as it is found both in the plural and the singular.

- (18) I'm waiting for your *feedbacks*, comments, insults and changes. (c, Italian, e-mail)
- (19) Please give your *feedback*! Thank you (b, Italian, e-mail)
- (20) Please let me know with *a feedback* (a, Italian, e-mail)

Tokens of this type appeared in the output of speakers from all three linguistic groups, which meant that the feature might have focussed. To determine this, we need to establish whether the use of the non-native form was affected by the same underlying factors in all three groups.

In this instance, there is no comparable native variability, so the patterns found in the non-native speakers are due to interference with their source languages or general problems of

acquisition. The analysis of this feature aims primarily to uncover whether there are factors which can be thought to influence the use of the non-native variant over the native form, as well as other factors which might inhibit the use of the non-native variant and whether these factors share the same constraints across language groups. This will help us establish whether the feature is involved in focussing.

As many of the lexical items appeared only infrequently in the corpus, it was decided to look solely at the term *information*, because unlike the other terms showing this type of variability, it occurred over 100 times in the corpus.

#### 4.3.2 Non-count nouns in English

English has a distinction between count nouns (which can be either singular or plural, examples 21-22) and non-count nouns (which have no clear plural, example 23). This distinction is relatively straightforward for non-native speakers of English to grasp, as it exists in a number of languages, including the source languages of our Swiss speakers, French, German and Italian.

(21) I have one dog.

(22) I have two dogs

(23) I have money/\*one money/\*two moneys

However, between languages, there can be a considerable amount of difference as to which nouns are considered count and which are considered non-count (*money* is plural in Greek, for example). As Quirk et al. (1985:251, emphasis mine) indicate ‘there is no necessary connection between the classes of nouns and the entities to which they refer. In some related languages, the nouns corresponding to **information**, money, news, and work, for example, are count nouns, but in English they are non-count’.

In English, *information* can only be used as a non-count noun (example 24), as attested by the fact that it can be used with *less* but not *fewer* (examples 25-26),<sup>52</sup> and because a ‘special’ construction is needed to represent it as singular (example 27). The difference can also be seen as a distinction between amount and number.

---

<sup>52</sup> This is irrespective of the fact that in current speech *less* is increasingly used with count nouns (i.e. there are less people here than yesterday).

- (24) \*I need some informations
- (25) He gave me less information than I needed
- (26) \*He gave me fewer information/s than I needed
- (27) He gave me a good piece of information

Although the use of *information* as a count noun in a plural form will not pose problems of comprehension to native or non-native speakers, native speakers will consider it to be non-native. The fact that it does not constitute a problem of comprehension may, in some part, explain why it is a feature used in the English spoken in Switzerland.

#### 4.3.3 Information in the source languages

The term *information* is used as a count noun in French, German and Italian, and can as such be used in both the singular and the plural forms (table 4.3). This means that the Swiss speakers' native categorization of the term will be at odds with the English categorization where *information* is used as a non-count noun.

French – une information, des informations
Italian – un'informazione, informazioni
German – ein Information, Informationen
vs.
English – a piece of information, information

Table 4.3: Singular and plural forms of *information* across languages

This difference in the distribution of *information* between the source languages and the target language, English, meant that it was unsurprising to find numerous cases of the form *informations* rather than the standard *information*, which were produced through interference with the structure in the source languages. Indeed, this distinction between count and non-count nouns is a known factor of interference in other non-native varieties of English (Erling, 2004:247). Erling (2004:247) found examples such as *She gave me an advice* and *I hope we don't have a lot of homeworks this semester* in her German university student data.

The interference is not constant, as the feature presents variation; there were cases of both variants in the speech of all three linguistic groups, and more importantly in the speech of the same speaker (see examples 28-33 below). Although the variation found in this feature is not entirely the type considered in sociolinguistic variation, the intraspeaker variation means that it can be used as a tool to establish whether the various groups share the same patterns. If it can be shown that the variation is not free but systematic, then we can try to establish whether the non-native groups share the same system. In a feature such as this one there are no native patterns to compare with the non-native groups.

Furthermore, as it should be assumed that the feature is interpreted as a count noun and as such can be used in the singular and the plural, it may be that although the unmarked forms of *information* appear to be what a native speaker may have used, the underlying structure of them is in fact that of a singular noun and not of a non-count noun. Evidence for this can be found in examples such as *an advice* above, where the singular form is overtly marked. This will have to be taken into account during the extraction and we will have to extract all the tokens of *information/s* and not just the ones with a plural function.

Examples (all e-mail)

Italian

(28) Please, send to me the missing *informations* requested in my report

(29) I am sorry so sorry.... I didn't know this *information* (both a – Italian speaker)

French

(30) So, it is the occasion to obtain some precious *information* about how to organise

(31) I must receive some more *informations* from NPO of Germany. (both f – French speaker)

German

(32) so i would like to collect already existing *informations* about the homepage

(33) add some other *information* about your country, university or anything else  
(both h – German speaker)

The presence of the *informations* variant reveals that this aspect was not taught in a way that inculcated Swiss speakers with the knowledge that it is not a Standard English form.

That the non-native speakers of English are, at times, unable to correctly use *information* as a non-count noun does not entail that they were not in fact taught how it should be appropriately used; non-native grammar and vocabulary books explicitly discuss the fact that there is a difference between languages and teachers stress to their students that although in French, German and Italian *information* is a count noun it is not in English.<sup>53</sup>

As *information* is used as a count noun in all three of the source languages, the underlying factors influencing the use of the two variants must be considered, as it is this which will enable us to disentangle the aspects which are due only to the speakers' source language from those which are due to focussing. We will be able to determine which are which by comparing the groups; aspects which are restricted to a single group are linked to the source language, while those which are found in all three groups might be due to focussing. This will make it feasible to establish whether the occurrence of the non-native form in all three groups is due to more than simple interference with the source languages. If the patterning of the three groups is similar across a range of factors then it is likely to be due to more than surface similarities.

Note that in features where it may help our understanding of the German speakers' patterns, both the High German and Swiss German forms are considered. Both of these varieties need to be examined, because the Swiss German students are in a situation of diglossia and speak a Swiss German dialect, but write in High German.

#### **4.3.4 Extraction and coding**

Before the analysis can take place, we need to establish which factors may constrain the variation found in the feature being considered. This means that it is very important to categorize (i.e. code) the data in such a way that will allow us to test the main hypotheses.

Each feature will have a different group of potentially important factors, of course, and these will be presented in the analysis of the actual feature. However, many of the external factor groups under scrutiny are relevant in all (or a large cross section) of the features selected for analysis and as such will be presented in depth here to avoid repeating the

---

<sup>53</sup> The teacher who taught my fellow students English definitely strove to emphasize this, 'a piece of information, not one information'.

reasons for their inclusion for each feature. The external factors which were systematically coded for in the various features are speaker, mother tongue and register.

### **External factors**

The external features tested in this analysis are somewhat different from those traditionally examined in variationist studies; as age, gender and social class have often revealed themselves to be very significant for variation patterns and are always integral to analysis (Milroy, 1987, Chambers, 1995, Labov, 2001). The present thesis does not make use of these three simply because they were not useful aspects to consider for the primary group considered in the analysis. All but one of the members of IFMSA are medical school students and as such they are all in their twenties and have had the same level of education. At the same time there are external factors which are pertinent to the study of non-native English which are not so for native English, namely mother tongue and individual speaker.

Although there are undoubtedly age differences in the English spoken in Switzerland, the present study only considered members of the medical students' association who all belonged to the same age group so it was not considered in the subsequent analyses.

Social class is not an altogether useful factor in the study of the Swiss non-native speakers, because they will have been exposed only to middle class English (i.e. Standard English). Although non-native grammar books do provide some exposure to different accents (via the tapes that are used in conjunction with them), they primarily regional differences (English, American, Scottish and so on) and not social ones. I will assume that the Swiss speakers all have the same standard forms of English as their target.

Gender was not considered as a factor for two reasons. First of all, none of the features selected in this thesis which had previously been studied for native English speakers have been found to be affected by gender. Poplack and Tagliamonte (1999) did not consider gender as a factor in their analysis of *going to*, nor was it considered in Tagliamonte, Smith and Lawrence's (2005) analysis of relative pronouns or in Tagliamonte and Smith's (2005) analysis of complementizers. Secondly, the non-native data was often distributed in such a way that male and female speakers had an unequal amount of tokens making any analysis of gender differences impossible.

### *Native language*

The speakers were divided according to native language so the patterns of the various linguistic groups could be analyzed separately. This was one of the most important factors as it allows us to establish what the patterns of the different linguistic groups were. Note that for the most part the French, German and Italian groups are presented separately.

### *Individual Speakers*

Speakers will be considered individually in each analysis to ensure that there are no idiosyncratic patterns skewing the overall distribution. The analyses will only present these marginals in the few instances where individual speakers were found to diverge from the rest of their linguistic group.<sup>54</sup> Previous studies (see for example Guy, 1980) have found that variability seen at a community level tends to apply to the individual as well; and this was largely the case of the features considered in this thesis as well.

Having discussed how the students are categorized, they need to be more fully introduced at this stage. Table 4.4 below provides information about the various speakers, their gender, mother tongue and the university they attended. Table 4.5 summarizes this information, showing the figures by gender and linguistic group.

Table 4.4: IFMSA-CH speakers

Speaker Code	Mother Tongue	University Language (if different)	Gender	University
a	Italian	g	f	BS
b	Italian	f	f	GE
c	Italian	g	m	BS
d	German		f	ZH
e	French		f	GE
f	French		m	GE
g	French		f	VD
h	German		m	ZH
j	French		m	GE
k	German	f	m	GE
l	French		m	GE

<sup>54</sup> Because of the uneven distribution of the data, it would have been confusing to present speakers individually for each feature.



m	German		f	BE
n	German		m	BE
o	German		m	BS
p	German		f	ZH
q	French		f	VD
r	German		f	BS
s	French		f	GE
t	German		f	BS
u	French		m	GE
v	German		m	BE
w	French		f	VD
T	Italian	g	f	BS
U	Italian	g	f	BS
V	Italian	g	f	BS
S	French		m	VD
R	French		f	VD
Q	Italian	f	m	VD
P	German		m	BS
N	German		m	ZH
M	French/English		f	VD
L	German		m	ZH
J	German		f	ZH
K	Italian	g	f	ZH
H	German		f	BE
*	French		f	GE
@	German	f	f	GE
%	French		f	GE
&	French		m	VD
#	German		f	ZH
\$	German		f	BS
~	German		f	ZH
-	French		m	GE
G	German		f	ZH
F	French		f	GE
E	German		f	BE
D	German		m	ZH
C	Italian	g	f	BS
B	German		m	ZH
A	German		m	ZH
l	Italian	g	M	BS

Table 4.5 Distribution of IFMSA-CH speakers by gender and mother tongue

	German	French	Italian
Female	13	10	7
Male	11	7	3
Total	24	17	10
% of speakers	47	33	20

### *Register*

The factor of register was very important to consider in this thesis; as mentioned in chapter 3, computer-mediated communication cannot be equated with speech, even in contexts where it is very informal. Because we have both speech and e-mail data for the non-native groups, we must establish whether register is affecting the patterning of the variable.

### **Coding of the lexical variable (*information*)**

All the tokens of *information*, in the standard form or the Swiss form, in the full form or the contracted form [information, informations, info, infos], were extracted from the IFMSA corpus.

The factor groups which this feature was coded for are somewhat more constrained than for some other features. This is a consequence of the fact that, being a lexical feature, there were considerably fewer tokens for this analysis than for some other features, and this made it more difficult to determine the importance of the various factor groups.

The external factors which this feature were coded for were: native language, speaker and register. The internal factors which were coded were the dependant variable (a marked plural form of *informations* vs. the unmarked form *information*), the contraction of the variants (information/s vs. info/s) and finally the type of determiner preceding the token.

## **4.3.5 Results and analysis**

### *Dependant Variable*

A total of 127 tokens were extracted from the Swiss data; 79 (58%) of them in the standard form *information* and 53 (42%) in the non-native form *informations*. This number of tokens will allow us to gain some insight into the factors which influence the distribution of variants. However, it is nonetheless a relatively low number, so some of the results will

need to be taken with special caution. This is because there may be cases of apparent patterns which are not only not significant but also potentially misleading.

The division in terms of register (table 4.6) is a good example why caution is important when considering features with low token numbers. There are far more tokens in the e-mail part of the corpus than in the oral part,<sup>55</sup> so we cannot determine whether the difference in percentage (43% vs. 25% usage of *informations*) is genuinely due to differences in use or simply because of the low number of tokens in the oral data.

Table 4.6: Distribution of variants across mode of communication in IFMSA corpus

	<i>Information</i>	<i>Informations</i>	Total
	%	%	N
E-mail	<b>57</b>	<b>43</b>	119
Oral	<b>75</b>	<b>25</b>	8
Total	<b>58</b>	<b>42</b>	127

#### *Native language*

Overall, the results demonstrate that the form *informations* is used 42% of the time, so while the standard form is used more, the non-native form is nonetheless used a considerable amount, especially bearing in mind that this is a form which the speakers would never have heard in native English speech. There is, however, a difference of usage in terms of language groups; table 4.7 reveals that while the Italian and the French speakers use the two variants almost equally (49% and 48% of *informations* usage respectively), the German speakers only use the non-native form a quarter of the time (26%). Although there are differences in the number of tokens for the three groups, this is not the cause of the difference in usage as a chi-square test does reveal that the distribution is significant.<sup>56</sup>

Table 4.7: Distribution of variations across languages in IFMSA corpus

	<i>Information</i>	<i>Informations</i>	Total
	%	%	N
Italian	<b>51</b>	<b>49</b>	65

<sup>55</sup> The e-mails make up the main part of the corpus so this is not unexpected.

<sup>56</sup> Chi-square = 6.00707804415762. *p* is less than or equal to 0.05. The distribution is significant. All the chi-square tests in this thesis were calculated using [http://www.georgetown.edu/faculty/ballc/webtools/web\\_chi.html](http://www.georgetown.edu/faculty/ballc/webtools/web_chi.html)

French	52	48	23
German	74	26	39
Total	58	42	127

The different rates of non-native use by the linguistic groups points towards the fact that the French, German and Italian speakers are not using the same variety of English, but we need to delve a little deeper into the features to determine whether the differences between languages may be influenced by other factors, such as individual speakers' use of the variants.

### *Individual Speakers*

Given the relatively low number of tokens, the results could have been skewed by the output of a single speaker; table 4.8 provides the distribution of tokens by language group for the speakers with more than 3 instances of the feature<sup>57</sup> and confirms this possibility as there are definite differences between speakers within the same linguistic group. The most striking is in terms of the German group, where, as table 4.8 reveals, there is only one speaker (h) who uses *informations*, whilst all the other German speakers follow the standard norm.

Table 4.8: Distribution of variants across speaker and main language

	<i>Information</i>	<i>Informations</i>	Total
	%	%	N
Italian			
a	12	88	8
b	70	30	41
c	13	87	15
1 other speaker	100	0	1 <sup>58</sup>
French			
f	55	45	11
g	0	100	3
7 other speakers	67	33	9 <sup>59</sup>
German			
A	100	0	4

<sup>57</sup> There are 19 different speakers who provided tokens of *information* or *informations*; only one token was obtained from 9 of them (50%) and two further speakers only provided two tokens.

<sup>58</sup> Chi-square = 20.604939447832, *p* is less than or equal to 0.001. The distribution is **significant**.

<sup>59</sup> Chi-square = 4.05509641873278. For significance at the .05 level, chi-square should be greater than or equal to 5.99. The distribution is **not significant**. *p* is less than or equal to 0.20.

h	62	38	26
m	100	0	3
5 other speakers	100	0	6 <sup>60</sup>

Although a chi-square test reveals that the results for the differences between the German speakers are not significant,<sup>61</sup> it does seem that with the exception of speaker h, the German speakers are using a very different strategy from the other linguistic groups. None of these German speakers use the *informations* variant. The possible reasons and implications of this will be discussed in the next section.

The results show differences in the Italian speakers as well, as there is a split between the two Italian/German speakers (87% *informations*) and the Italian/French speaker (30% *informations*). This may be due to the influence of French and German, but as it is the Italian/German speakers who make the most use of *informations* and the German speakers appear, except for one speaker, not to use it at all, this is unlikely. Furthermore, while the German distribution was not statistically significant, this one is. Again, the low number of tokens makes it difficult to establish how important these findings are.

These results suggest that the variation found in this feature is largely personal rather than linked to native tongue. Although this may be linked to the English proficiency of the individual IFMSA-CH members, remember that the members have all had similar amount of English language instruction.

The individual effect in the use of *informations* was substantiated to some extent by the results found in the other Swiss data collected for the research project. There were no instances whatsoever of the non-standard feature in any of the other corpora (table 4.9). This fact, however, was undoubtedly linked to the overall low number of tokens of *information* in these corpora (22 tokens across the three corpora).

Table 4.9: Distribution of *informations* across data sources

	<i>Information</i>	<i>Informations</i>	Total
	%	%	N
IFMSA	58	42	127

<sup>60</sup> Chi-square = 6.72413793103448. For significance at the .05 level, chi-square should be greater than or equal to 7.82. The distribution is **not significant**. *p* is less than or equal to 0.10.

<sup>61</sup> Note that the low number of tokens per cell means that these results need to be viewed with caution.

Other project data (First Tuesday, M4Music and bank interviews)	100	0	9
Total	<b>64</b>	<b>36</b>	149

These results should not be taken to mean that *informations* is not found in other realms of English use in Switzerland than IFMSA; indeed by using a Google search solely on Swiss pages in English, numerous other instances of Swiss English users producing *informations* alongside *information* can be found (see examples 34-35 below).

- (34) You can find informations in english in the "Downloads/Presse" section <sup>62</sup>  
 (35) Informations about state of Horus Networks Sàrl services. <sup>63</sup>

The considerable differences in variable selection for individual speakers, coupled with the fact that the three linguistic groups are not following the same strategies, point towards a conclusion that this feature is not currently undergoing focussing. Each group, or more accurately each individual, is following a distinct strategy.

It is also interesting that French speakers make use of the plural *informations* form at all, as elsewhere there were numerous occurrences of missing plural marking in their output (examples 36-37 below).

- (36) ‘we need more doctors but the problem is that the cantonØ don’t  
 want to pay more for health’ (w – French speaker, interview)  
 (37) we were together in an English school for two monthØ and in two  
 monthØ we never spoke French together (S – French speaker, interview)

This is due to the fact that in French, there is no phonological distinction between plural and singular on the noun itself (le chat, les chats [ləʃat] vs. [ləʃat]). This feature was not selected for analysis as it was restricted solely to the French speakers and could not provide insight into processes of focussing. However, it means that the occurrences of *informations* as an overly marked plural noun are all the more significant, as it underlines the extent to which *information* is seen by French speakers as a noun which requires plural marking.

<sup>62</sup> Retrieved September 5, 2006 from: <http://twiketours.com/challenge/content/view/29/76/>

<sup>63</sup> Retrieved September 5, 2006 from: <http://www.horus.ch/clients-informations.en.shtml>

Despite the fact that this feature is not in the process of focussing across all three groups, it is worthwhile to continue the analysis of this feature to determine if there are factors influencing the choice of variants for the speakers who do use more than one variant. This may allow us to establish whether there are other similarities in the distribution of the feature in these speakers.

### *Preceding Determiner*

A factor which may prove to be significant for the choice of variant is the determiner preceding *information* or *informations*. The non-native speakers' choice of form may have been influenced by the type of determiner; e.g. *more* might have triggered a plural form more often than *this* as it is more commonly associated with plural forms. These would be cases where the Swiss languages would be most likely to have the overtly pluralized form. Table 4.10 presents the breakdown of the *information* tokens according to the determiner preceding them and in terms of the IFMSA corpus as a whole. The factor verb + *info* comprises structures such as the one in example 38. Note that all linguistic groups are considered together as the numbers in each cell would have been too low otherwise.

- (38) 'This is a database which provides information and facts about the medical curricula' (o – German, e-mail).

Table 4.10 Distribution of variants according to the determiner in the IFMSA corpus

	Information	Informations	Total
Determiner	%	%	Ns
The information	58	42	26
More information	35	<b>65</b>	17
Some information	<b>64</b>	36	25
Further information	50	50	6
This information	<b>86</b>	14	7
All the information	60	40	10
No preceding det.	60	40	10
Verb + info	<b>86</b>	14	7
Other (any, no, these, etc.)	53	47	19
Total	58	42	127

While the preceding determiner influences the choice of *information* versus *informations* and vice versa to some extent, the number of tokens is too low to be able to determine precisely how much this is the case; too many of the factors only have a handful of tokens and thus their results cannot be considered to be definitive. The determiner *more* influences

the use of the plural variant, however, as out of 17 tokens, 65% appear as *informations* which is far more than the overall average of 42%. No general patterns of use can be uncovered however, as it would be expected that *some* and *more* would have patterned in a similar fashion and they do not. This also meant that it was not feasible to divide the results by pluralizing determiner vs. other determiners as this would have skewed the data.

Also, *this* is found with *informations* in one instance (example 39); this use of the singular determiner in place of the plural one was found throughout the corpus and is linked to the non-natives' pronunciation of the vowels /ɪ/ and /i:/ in *this* and *these*, which can be very similar (example 40).

(39) 'well that's our job to deliver **this infos**.' (h, German, e-mail)

(40) Please, make possible that we have **this** places. (c, Italian, e-mail)

#### *Full vs. contracted form*

An examination of the use of the full forms (*information* and *informations*) as opposed to the contracted forms (*info* and *infos*) could reveal differences in variant selection, as well as help to determine whether the three groups share the same processes. Furthermore, this factor group is also relevant in native speaker English, and this will enable a comparison between the native and non-native groups (table 4.11). Of course, for the non-native speakers there are four possibilities to the native speakers two, but the full and contracted form distinction remains.

The native speaker corpus did not provide a very large number of tokens; there were twelve in the whole corpus. This will nonetheless allow us to form an idea of the relative frequency of *info* vs. *information* in the corpus.

Table 4.11: Distribution of full and contracted forms of *information* for native and non-native speakers

	Information	Info	Total
	%	%	N
French			
Plural form	100	0	11
Non-count/singular form	75	25	12
Total	87	13	23
German			
Plural form	60	40	10



Non-count/singular form	97	3	29
Total	87	13	39
Italian			
Plural form	72	28	32
Non-count/singular form	82	18	33
Total	77	23	65
Overall non-native			
Plural form	75	25	53
Non-count/singular form	86	14	74
Total	82	18	127
Native			
Non-count form	64	36	42

Although separating the non-native speakers into their appropriate linguistic groups makes the number of tokens in each cell rather small, it is a necessary division as table 4.11 demonstrates that there are considerable differences in the three groups. While the contracted form severely restricts the use of the supposedly Swiss form for the French speakers, it is quite the opposite for the German and the Italian speakers who are more likely to produce *infos* than *info*.

The use of the contracted form is lower in all three non-native groups than it is for the native speakers (13 to 18% opposed to 36%), even when the standard and the non-native forms are considered together.

#### *Multivariate Analysis*

Unfortunately, the low number of tokens produced in the three groups means that a multivariate analysis cannot be conducted for this feature.

### **4.3.6 Discussion**

The results found in this analysis tend towards the conclusion that the speakers of the three source languages do not share the same strategies, as the overall distribution in terms of speaker groups and individuals does not reveal any general tendencies.

Furthermore, the results reveal that the differences are linked more to idiosyncrasies of the individual speakers than to native language; this is linked to difficulties in examining

features of interference-based variation as the level of competence of the speaker can play an important role.

Although *informations* can be considered a Swiss feature, because it is found in all three English speaking groups in Switzerland, it cannot be considered to be indicative of focussing. Table 4.12 below, presents an overview of the findings for this feature.

Table 4.12 Overview of differences between non-native speaker groups

	French	Italian	German
Overall distribution of <i>informations</i> form	Like ITL	Like FRE	Unlike others
Clear inter-speaker variation	No	Yes	Yes
Preceding determiner similarities	Yes	Yes	Yes
Contraction similar to natives	No	No	No

In terms of overall distributions, the French and Italian speakers are quite similar (though not identical), while the German speakers use far less of the Swiss variant. The three groups do not share the same patterns. The fact that the analysis found inter-speaker variation in the Italian and German groups adds a further dimension to this; this variable is perhaps more influenced by speaker than by source language. This is a case where the native language of the speaker is the main factor in the use of the non-native variant (case C).

Case C	French, German and Italian vary unlike native speakers and unlike each other	variation influenced by native language only
--------	--	--

## Part II

### 5. Future

#### 5.1 Introduction

Unlike some other languages, French and Italian among them, English has no specific future tense but instead makes use of other syntactic strategies to convey future time (Bybee et al., 1991). Indeed, in Standard English ‘futurity, modality and aspect are closely interrelated, and this is reflected in the fact that future time is rendered by means of modal auxiliaries, by semi-auxiliaries, or by the simple present or present progressive forms’ (Quirk et al., 1985:213).

Native speakers are often able to choose from several possible constructions to convey futurity. Berglund (1997), Poplack and Tagliamonte (1999) and others have studied the distribution of the various future forms and have demonstrated that the selection of these forms is constrained in quite specific ways by a range of factors. Furthermore, they established that there are a number of cross-dialectal and cross-speaker similarities in terms of factors conditioning the variability. Consequently, some of the rules governing the distribution of future forms can be said to be shared to some extent by all native speakers of English. Thus, this variable provides an excellent situation to compare native to non-native variability. In addition to this, the two main future forms *will* and *going to* are presented more or less as variants of each other when they are taught in Switzerland, providing us with a case where we can establish whether the overt teaching of the variants helped the acquisition of native patterns.

This chapter will consider the factors found to be significant in the abovementioned studies and attempt to establish whether the non-native speakers of English in Switzerland follow native constraints for variant selection.

#### 5.2 Future Forms

### **Will future**

The modal auxiliary construction is ‘the most common way of expressing futurity’ (Quirk et al., 1985:214), and uses *will*, *shall* or *’ll* (examples 1-3).<sup>64</sup> The form *will* can be used with all persons, while ‘the form *shall* can only be used with a first person subject’ (Quirk et al., 1985:214). This is linked to ‘a strong teaching tradition [which] has upheld the use of *shall* as the correct form, in preference to *will*, with a first person subject in formal style’ (Quirk et al., 1985:214).<sup>65</sup>

- (1) Our next national meeting **will** take place (e, French, e-mail)
- (2) **Shall** I send a copy of it to each LEO? (c, Italian, e-mail)
- (3) I**’ll** be quiet. (h, German, e-mail)

The various modal forms are said to approximate ‘a colourless, neutral future’, generally associated with prediction (Quirk et al., 1985:214).

### **Will + progressive infinitive**

This construction ‘combines reference to a future time with the ‘temporal frame’ associated with the progressive’ (Quirk et al., 1985:214) (example 4).

- (4) I **will be stressing** you (c, Italian, e-mail)

### **Going to future**

*Be going to* followed by the infinitive is ‘frequently used to express futurity, especially in informal speech’ (Quirk et al., 1985:214) (example 5). Its general meaning is ‘future fulfilment of the present’, but it can also be used to express ‘future fulfilment of present intention’<sup>66</sup> and ‘future result of present cause’.<sup>67</sup>

- (5) I don't think she **is going to** come in Lausanne. (S, French, e-mail)

### **Simple present future**

For Quirk et al. (1985:215), ‘the simple present is, after the *will/shall* construction, the **next most common** means of referring to future actions in English’ (emphasis mine). It should be noted, however, that ‘this future use of the simple present is frequent, however, **only in dependent clauses**, where it is regularly used after condition and temporal conjunctions

<sup>64</sup> Although Quirk et al. provide judgments about the frequency of the various forms, their work does not provide a quantitative study of this.

<sup>65</sup> *Shall* is also used in colloquial English in 1<sup>st</sup> person questions consulting the needs of the listener (as in example 2) (Quirk et al., 1985:230).

<sup>66</sup> ‘chiefly associated with personal subjects and agentive verbs’ (Quirk et al., 1985:214).

<sup>67</sup> ‘found both with personal and nonpersonal subjects’ (Quirk et al., 1985:214).

such as *if* and *when* as well as in some *that*-clauses' (Quirk et al., 1985:215, emphasis mine).

(6) Tomorrow (7th of April) **is** the World Health Day (g, French, e-mail)

(7) If I **don't get** answer from you beautiful people, I'll start calling you home at every hour of the night (c, Italian, e-mail)

### Other future forms

Quirk et al. (1985) discuss further future forms such as the progressive present, and the *be to* forms which also convey futurity. These appear less frequently than the other forms however.

Quirk et al.'s judgments about the relative frequency of the various future forms are corroborated by empirical data. Poplack and Tagliamonte (1999:326) found that in Ottawa, the most mainstream of the varieties they considered, the *will* and the *going to* forms made up 74% of the overall distribution, while the progressive present and the simple present future together accounted for only 26%.

Although the various constructions for the future tense can all be found in non-native English, as demonstrated in the examples above, this study will only focus on the two most common future constructions, the *will* future and the *going to* future,<sup>68</sup> leaving aside the *simple present* future forms as they are mostly restricted to dependant clauses. Looking solely at these two variants will enable a more direct comparison with the results of both Poplack and Tagliamonte (1999) and Berglund (1997) as these are the variants which were examined most closely in their studies as well.

Poplack and Tagliamonte's research also focused on the diachronic grammatical change associated with the future forms, using it to account for some of the patterns they found in their data; these aspects will not be considered in the present analysis, however, as our aim is to establish whether the non-native speakers are using the same patterns as the native speakers rather than to account for the source of these patterns.

Although these two ways of expressing futurity diverge semantically to some extent,

---

<sup>68</sup> *Will* + progressive is considered to be an aspect of the standard *will* future form.

‘there are occasions where the choice of one construction (say *will* + infinitive) rather than another (say *be going to* + infinitive) has scarcely perceptible effect on meaning’ (Quirk et al., 1985:213).

The notion that there is a degree of choice between the two variants is discussed in language teaching literature as well; the ‘choice [between the two forms] depends on how the speaker sees the event and not on its certainty or nearness to the present’ (Soars & Soars, 1987:54).

Examining how the two variants are taught will help us understand how non-native speakers of English in Switzerland have been introduced to the variable. Despite some discussion that the forms can be used interchangeably (as in the quotation above), for the most part the English grammar books used in Switzerland present a fairly prescriptive view of the forms. Students are taught that *will* is used for prediction and *going to* is used for ‘future fulfilment of the present’ (Quirk et al., 1985:214). Students are taught (and tested on) when it is appropriate to use one form and when it is appropriate to use the other. Note, however, that the teaching of the future forms does not focus on the notion that *going to* is the less formal form; the variants are presented as being used in specific situations but not in terms of formality. This is an important point as Berglund (1997) had found that register (oral vs. written data) and formality did affect the use of *going to*.

This fact will enable us to determine if non-native speakers acquire variability which is explicitly taught. Although, as mentioned, notions of prediction or fulfilment of the present can be difficult to disambiguate, they are linked to factors which have been associated with the variability in the future in native speakers to some extent.

If we accept that variability may be a difficult feature for non-native speakers to acquire, this chapter will help to establish whether the formal teaching of variation can contribute to the acquisition of variable patterns. Will the fact that the Swiss speakers have been explicitly taught when to use *will* and when to use *going to* mean that their production of the two variants will match that of native speakers’?

On the surface, all three Swiss groups use both future forms in their English output, as is demonstrated in the examples below.

All the examples below are from e-mails

French speakers:

- (8) *I'll see* with Basel who *is going to do* the inscription (\*, French)
- (9) I *will* also come tomorrow to the meeting in ZH. (s, French)
- (10) Well, I hope that a solution *will* be found! (\*, French)

German speakers:

- (11) I would really appreciate if you could prepare yourself for your standing committee *your gonna* join. (h, German)
- (12) then we go together to the hospital where the meeting *is going to take* place (h, German)
- (13) I *will* come on saturday not having decided which train yet but soon enough! unfortunately *I'll have to leave* on (p, German)
- (14) About the Schoki, I *will* buy the 10 Kg and take them (o, German)

Italian speakers:

- (15) Once back, they *won't* never ask again the sense of their volunteer job. (c, Italian)
- (16) I *will* also try the pharmaceutical company "X". (V, Italian)
- (17) *I'll* also contact our social programme group. (V, Italian)
- (18) Let me know who else *is gonna come*!? (P, Italian)
- (19) I received today the official program. *Shall* I send a copy of it to each LEO? (c, Italian)

### 5.3 Future forms in the three Swiss languages

The way the future meaning is encoded in the three native languages needs to be considered, as this may help us identify any non-native English patterns which are different from the native norm at a later stage of the analysis.

*Will* form

Unlike English, French and Italian have a future tense. This tense is marked inflectionally on the verb itself; each person has a separate morphological future form tagged onto the verb root (examples 20-23). Note that the opposition between inflectionally rich verb forms in French and Italian and uninflected verb forms in English is a general pattern.

#### French

(20) Demain j'**irai** à la plage

(21) Demain tu **iras** à la plage

#### Italian

(22) Domani and**rò** alla spiaggia

(23) Domani and**rai** alla spiaggia

German syntax does have some inflectional morphology, but, unlike French and Italian, it does not have a separate construction for the future. Instead, similarly to English, one of the forms to express futurity is the use of a modal verb (*werden*). This is not, however, a very frequent form in spoken German and 'in practice such forms are rather restricted in use' (Fox, 1990:183). A similar form exists in Swiss German as well, but as in the case of High German, the *will* construction is not frequently used.

(24) Morgen werde ich an den Strand gehen

#### *Going to* form

Italian and High German do not have a future form comparable to the English *going to*. French speakers, on the other hand, make use of a construction similar to the English one (Bybee et al, 1991). This construction involves the verb *aller* (to go) and an infinitive, allowing a French speaker to say, for example, 'je vais manger' as well as 'je mangerai'. The factors influencing the use of the *going to* form in French are similar to those of English (see Emirkanian and Sankoff, 1985), as they share a 'future fulfilment of present situation' sense as discussed by Quirk et al. (1985:214).<sup>69</sup>

Moreover, the *aller* + inf is the main form in informal French at present, and studies of Canadian French (Emirkanian and Sankoff, 1985:194, Poplack and Dion, 2004) revealed

<sup>69</sup> Compare for example the different meanings of 'She will have a baby' and 'She's going to have a baby' to 'Elle aura un enfant' and 'Elle va avoir un enfant'.



that it was used nearly 80% of the time, far more than the inflectional form. The use of the *aller* form is said to be quite high in continental French (including Swiss French) as well (Ball, 2000:87). The Swiss German dialects also have a similar construction ‘I gan go esse’ (I going to go eat).

#### Other future forms

The use of the simple present to convey the future is available in the three Swiss languages as well. Its function in French, German and Italian is similar to that in English. Unlike English, however, it is not restricted to dependant clauses. This form is the one most commonly used in German (Fox, 1990:183).

Table 5.1: Comparison of future structures

Future forms	Suffixation	Modal verb	‘ <i>Going</i> ’ based form
English		√	√
French	√		√
Italian	√		
German		√	
Swiss German		√	√

## 5.4 Previous studies

Previous studies on the future (Berglund, 1997, Poplack & Tagliamonte, 1999 and Tagliamonte, 1997, 2003) have determined that the constraints on the choice of variant are style, register (oral vs. written speech) (Berglund, 1997), the point of reference, age of the speaker and personal pronoun used (Poplack and Tagliamonte, 1999). These studies found varying rates of *going to* use, ranging from 21% in Berglund’s analysis of the London Lund Corpus, to 30% in Tagliamonte’s York corpus, to 45% in Poplack and Tagliamonte’s Ottawa corpus. These are the figures which this analysis will use as a benchmark for the non-native results. Poplack and Tagliamonte (1999) found that the internal factors which constrain the use of the variables were cross-dialectally shared for the most part; so British and North American varieties of English had very similar hierarchies and ranges in their variable rules for the selection of *will* and *going to*.

## 5.5 Data sources

For the analysis, the future variants selected for analysis were extracted in the whole of the IFMSA corpus (both e-mail and oral data) and in the native speaker e-mail corpus. These non-native oral results will also be compared to those of other studies of native speaker future use.

## 5.6 Extraction and Coding

All the tokens of *will* and *going to* as future forms were extracted from the corpora. Tokens of *going to* that only conveyed directional meanings were excluded (examples 25-26).

(25) All the people who *are going to Denmark* are kindly invited to a big party  
(h, German e-mail).

(26) just wanted to say that *I am going to Basel* on 4-5th May. (S, French, e-mail)

The tokens were coded for variant (*will*, *going to* and *shall* as a sub-variant of *will*), for grammatical person, for whether the form was contracted (*'ll* or *gonna* vs. *will* and *going to*<sup>70</sup>) and, in the Swiss data, speaker identity, speaker native language and register. The contraction of the form is linked to orthographic convention, of course, and was not considered in the oral data. It was chosen as a factor because it can help us establish the formality of the e-mails. Grammatical person was considered for two reasons; first of all, Poplack and Tagliamonte (1999:335) had found that *going to* occurred with less first person subjects than other subjects. Secondly, because we are dealing with both oral and e-mail data, this factor will help us establish if different subjects occur at different rates in these two registers. This may help us better understand the results.

The non-native tokens were also given an additional coding if they were felt to be in some way different from native-speaker English (examples 27-28). This was to allow us to analyze the interference-based variability found in this feature. Section 5.7.2 will focus on the reasons why the tokens were judged to be non-native.

---

<sup>70</sup> By contracted form for *going to* the contraction between *going* + *to* is meant rather than the subject with the verb to be.

- (27) A booklet that will include all the information I ***will*** get from you (b, Italian, e-mail)
- (28) They ***shall*** say the swiss have good organized hospitals on a high level (h, German, e-mail)

Fewer factor groups were considered in this study than in previous research, this is due to the nature of the variable in the Swiss speakers' English as will become clear below.

## 5.7 Results

### Overall distribution

The extraction yielded a large number of tokens; nearly 1500 tokens of e-mail and oral data combined for the three non-native groups, and a further 237 tokens from the native e-mail corpus.

Tables 5.2 and 5.3 below give the overall distribution of *will* and *going to* by non-native groups for e-mail and oral data respectively.

Table 5.2. Percentages of *will* vs. *going to* in non-native speaker e-mails

	% of <i>will</i>	% of <i>going to</i>	Number of <i>going to</i> tokens	Total N
Native language				
Italian	<b>99</b>	<b>1</b>	7	650
French	<b>99</b>	<b>2</b>	7	314
German	<b>93.5</b>	<b>6.5</b>	13	201 <sup>71</sup>
<b>Total</b>	<b>98</b>	<b>2</b>	<b>27</b>	<b>1165</b>

Table 5.3. Percentages of *will* vs. *going to* in non-native speech

	% of <i>will</i>	% of <i>going to</i>	Number of <i>going to</i> tokens	Total N
Native language				
Italian	<b>99</b>	<b>1</b>	<b>1</b>	186
French	<b>95</b>	<b>5</b>	<b>3</b>	64
German	<b>65</b>	<b>35</b>	<b>7</b>	20
<b>Total</b>	<b>96</b>	<b>4</b>	<b>11</b>	<b>270<sup>72</sup></b>

<sup>71</sup> Chi-square = 19.7402324545435. *p* is less than or equal to 0.001. There is a significant difference across language groups.

<sup>72</sup> The low numbers for *going to* in three groups make it impossible to use chi square to determine if the distribution between the three groups is significant.

These tables reveal that the rates of *going to* are low for all groups in both the oral and the e-mail data. They also show that the German speakers have significantly higher rates of *going to* use than the other groups in the e-mail data. Note, however, that although there very few *going to* tokens in the data, the tokens are not due to the idiosyncrasies of individual speakers in each linguistic group. In terms of the e-mail data, the distribution found in each linguistic group is due most of the speakers using *going to* at least once.<sup>73</sup> From this we can conclude that *going to* is in the linguistic repertoire of all the speakers but *will* is used at very high rates by all of them.<sup>74</sup>

The results of the oral data must be taken with more caution, in terms of the German speakers especially. The German distribution consists in only 20 tokens and all of the *going to* tokens in the German oral data stem from a single speaker (who uses both *will* and *going to*), whilst the other German speakers use only the *will* form. Except for this one speaker then, the German speakers are quite similar to their French and Italian counterparts. All three groups have very low rates of *going to* in oral and e-mail data. This is an important finding as previous studies (Berglund, 1997:15) had found that register had a significant effect on *going to*.

Tables 5.4 and 5.5 below present the overall distribution of *going to* in native-speaker studies. Table 5.4 provides figures from Poplack and Tagliamonte (1999), Tagliamonte (1997) and Berglund (1997), while Table 5.5 provides the figures from the native e-mail data collected for this thesis.

Table 5.4: Overall distribution of *will* and *going to* in earlier studies (oral data)

	% of <i>will</i>	% of <i>going to</i>
Ottawa (Canadian English)	54	46
York (Northern British English)	70	30
London (LLC) (Southern British English)	79	21 <sup>75</sup>

Table 5.5. Percentages of *will* vs. *going to* in native speaker e-mails

	% of <i>will</i>	% of <i>going to</i>	Total N
Native English Total	<b>89.5</b>	<b>10.5</b>	237

<sup>73</sup> This represents 100% of the Italian speakers, 60% of the German speakers and 71% of the French speakers once individuals with fewer than ten tokens had been removed.

<sup>74</sup> Some of these speakers had fewer than one instance of *going to* in 200 tokens

<sup>75</sup> The figures for Ottawa come from Poplack and Tagliamonte (1999:326), those for York from Tagliamonte (1997) and those for London are from Berglund (1997:15).

The figures from the e-mail data are lower than those for oral data; this was the predicted outcome, as *going to* has been found to occur less in writing and e-mail than in speech. Although Berglund (1997:15) found 21% of *going to* in the oral data, the written corpora she considered (the Brown Corpus and the LOB) had significantly lower rates (around 5%). In the case of *will* vs. *going to*, e-mails are much closer to writing than speech.

How does this compare to the non-native data? The non-native e-mail results are very low in terms of *going to*, but unlike the native data, there is no shift towards higher rates of *going to* in speech. Removing the one German speaker who had very different frequencies of use, the overall distribution of *going to* in the oral data remains very low in all three non-native groups. Figure 5.1 presents the oral/e-mail distribution across the four linguistic groups.<sup>76</sup>

**Figure 5.1: % of *going to* in oral vs. e-mail data**

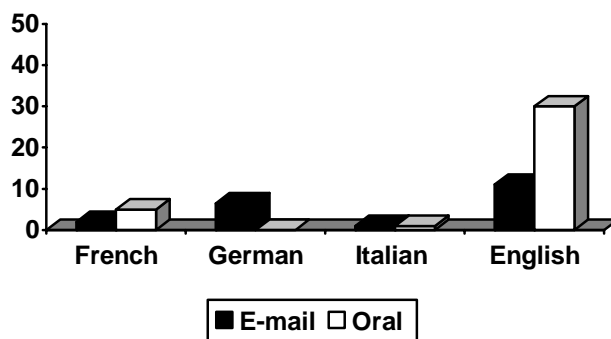


Figure 5.1 confirms two points; first of all, that the non-native rates are consistently lower than the English rates. Secondly, that the non-native speakers do not have the register distinction of native speakers. Although there is a slight increase of *going to* in oral contexts for the French speakers, the difference is not statistically significant<sup>77</sup>. The three non-native groups show the same pattern.

<sup>76</sup> The oral English figures come from the York speakers in Tagliamonte (1997) which represents the group most similar to the e-mail corpus.

<sup>77</sup> Chi-square = 1.24734978624342. *p* is less than or equal to 1.

These results underline the importance of using a comparable corpus of native e-mails with the non-native data; it has allowed us to establish that the difference between natives and non-natives is not due solely to register. Although it is plausible that the non-native e-mailers are more formal in their writing than the native e-mailers, the results do not lend themselves to this conclusion, as the discrepancy between native and non-native patterns is maintained in speech.

### Grammatical Person

The near categoricity of *will* use means that a further analysis into many of the factors which were found to condition variation in previous studies could not be conducted. One factor examined in previous studies which we can consider is grammatical person. Poplack and Tagliamonte (1999:335) found that, although historically *going to* was linked to speaker attitude and hence was favoured in first person contexts, in most contemporary varieties of English *going to* had become grammaticalized to the point that it occurred more often with non-first person subjects. Table 5.6 examines the four linguistic groups in terms of the e-mail data. The non-native oral data was not included in this analysis as my results could not be compared with those of previous studies<sup>78</sup> and because of a low number of *going to* tokens in non-native speech. The proportion of first person/non-first person tokens was very similar for all groups across register, however, averaging around 40%.

Table 5.6 Percentage of *going to* by grammatical person

	1st person (sing and plural)		Non-first person	
	% of <i>going to</i>	Total Ns	% of <i>going to</i>	Total Ns
Italian	1	309	1.2	341
German	1.1	88	<b>10.6</b>	113
French	2.4	126	2.1	188
English	8	76	12	161

There is no difference in the distribution of *will* and *going to* in terms of grammatical person for the Italian and the French speakers, their rates of *going to* remain very low regardless of person. This may be an effect of the extremely low rate of *going to* tokens overall. The English and German speakers, however, follow the predicted pattern in that they have higher rates of *going to* in non-first person contexts. The German pattern is especially plain, with a rate of 1% in first person contexts going up to nearly 11% in non-

<sup>78</sup> Poplack and Tagliamonte (1999) had provided the factor weights for grammatical person but not percentages.

first person contexts. This is backed up by the fact that the difference was found to be significant for the German speakers,<sup>79</sup> but not for the English e-mailers. The German speakers have acquired the native English constraint uncovered by Poplack and Tagliamonte (1999:335), but the Italian and French speakers have not.

### Full vs. contracted forms

A factor that may reveal further similarities or differences between the groups is whether the verb forms *will* and *going to* are contracted to *'ll* and *gonna* respectively. The contraction of the two variants functions in slightly different ways so we will consider full and contracted forms of the two variants separately. Although this factor cannot help us learn more about the selection of one of the variants over the other, it does offer us the opportunity to examine the effect of the e-mail register. Table 5.7 presents the contraction in the native e-mails, while Table 5.8 presents the non-native e-mails.

Table 5.7. Percentages of full forms vs. contracted forms in native speaker e-mails

Native English	% of full form	% of contracted form	Total N
<i>Will</i>	77	23	212
<i>Going to</i>	92	8	25

Table 5.8. Full and contracted future forms in non-native English e-mails

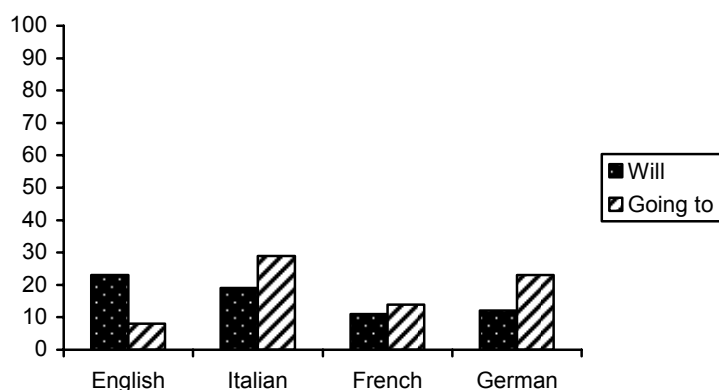
	% of full form	% of contracted form	Total N
Italian			
<i>Will</i>	81	19	643
<i>Going to</i>	71	29	7
French			
<i>Will</i>	89	11	307
<i>Going to</i>	86	14	7
German			
<i>Will</i>	88	12	188
<i>Going to</i>	77	23	13

The native speaker rates of contraction are 23% for *will* and 8% for *going to*; this is fairly low and underlines how even messages written in an informal style are affected by the register of e-mail. Turning to the non-native groups, we can see that they differ considerably from the native group, although they are quite similar to each other. Their rates of *will* contraction are slightly lower than in natives, but their rates of *going to*

<sup>79</sup> Chi-square = 7.2534847562222. *p* is less than or equal to 0.01.

contraction are higher. The rates for *going to* need to be viewed with caution, however, as there were very few *going to* tokens in the non-native data. Figure 5.2 compares the percentages of *will* and *going to* contraction across language groups.

**Figure 5.2: Percentage of contraction**



The high use of *gonna* by the non-native speakers in e-mails may show to what extent it is used as an idiomatic form in the English of the Swiss speakers. Although *gonna* is found in informal contexts in written native English as well, this does not account for the non-native use on two counts. First of all, if this had been the case then we would have expected *will* to be contracted more often as well. Secondly, the proportion of *going to* increases in informal contexts in native speech; if *gonna* were linked to informality for the non-native speakers, then the overall use of *going to* forms would be higher.

Although only the German speakers had shown a significant difference between the first person and non-first person subjects in terms of the distribution of *will* and *going to*, this factor also needs to be considered for the contraction of *will*. If first and non-first person subjects do not contract in the same way across language groups then this could have an impact on our results as Table 5.6 had revealed that nearly half the future tokens in the data were first person. Table 5.9 considers this, examining the contraction of *will* in first person singular and non-first person contexts across the linguistic groups.

**Table 5.9: Contracted forms of *will* by grammatical person**

	1st person		Other	
	% of contraction	Total N	% of contraction	Total N



French	20	123	4	184
German	17	87	7	101
Italian	29	306	10	337
English	41	70	14	142

Table 5.9 shows that all groups are more likely to contract *will* with first person singular subjects. This distribution was found to be statistically significant in all four groups. What is especially important for this analysis is the fact that the non-native speakers show lower rates of contraction in all contexts, although the direction of effect is the same.

### Interference-based variation

In addition to comparisons with the variation found in native speakers, this feature also can offer insight into interference-based variation, allowing us to study whether the non-native groups share the same patterns when using the various future forms.

Due to the overall low use of *going to* in the non-native data set, most of the tokens which were judged to be non-native are of *will*. Generally, this involves cases where *will* is overused. About 8% of the tokens of *will* did not follow native speaker norms (table 5.10). Note that these are not cases where *going to* should have been used. The sentences were found to be non-native for a number of reasons discussed below.

Table 5.10: Number and percentage of non-native sentences in three language groups

	French	German	Italian	Total
Total number of non-native sentences	20	16	59	95
Percentage of non-native sentences	6.3	7.9	9.1	8.1

### Will in place of a continuous form

There were a number of cases where *will* was used in situations where Standard English uses a present (or future) continuous tense. This overuse of *will* was especially common when it was with the verb *to come*.

- (29) Just let us know how many of you ***will come***, so that we buy enough food (c, Italian, e-mail)
- (30) unfortunately I ***won't*** meet you in Basel...because I ***will stay*** in New York at that time (d, German, e-mail)

### ***Will for would***

Some of the interference-based tokens found in the data were caused by the speakers using *will* instead of *would*. Standard English requires a conditional tense rather than a future tense in sentences such as those in the examples 31 to 33 below.

- (31) I am leaving for Malta on the 2nd March, so it ***will*** be nice of you, if you could send me these reports before Thursday (g, French, e-mail)
- (32) In all cases you will have to send to me the REPORT of the activities of this past 6 months. I ***will*** be extremely grateful if you could sent them AGAIN to me, even if you sent them out already (b, Italian, e-mail, emphasis in original)
- (33) the discussion and recruitment of new people ***shall*** start already now, and not during the Presentation of the Kopaonik Final Report, or during the SCORE Session (c, Italian, e-mail)<sup>80</sup>

### **Tense concord**

Interference-based variation was commonly caused by tense concord (examples 32-35), namely cases where the tense agreement rules followed were those of the native language rather than English. In conjoined future tense clauses, French, Italian and German tend to have a future form in both the main and the subordinate clause, whereas English has a future tense only in the main clause and a simple present in the subordinate clause (as mentioned in section 5.2).

- (34) This summer, when the intranet section ***will*** be available, an anti-virus ***will*** maybe be in there for download (j, French, e-mail)
- (35) All in all it seemed to be a good appointment and I ***will*** let you more about the story as soon as I ***will*** know more, alright? (d, German, e-mail)
- (36) As soon as ***I'll*** get them I ***will*** send them to you all (b, Italian, e-mail)
- (37) I can tell from Geneva that many thing ***will*** be decided when we ***will*** meet the vice-dean who (b, Italian, e-mail)

### ***Shall***

---

<sup>80</sup> Actually in this case *shall* is used in place of *should* which is slightly different from the other examples.

*Shall* was occasionally used with pronouns other than *I* and *we*. This may in fact be a case of hypercorrection rather than language interference, in that the non-native speakers are using a form which they have been taught as being more appropriate than *will*. Non-native English grammars try to enforce the use of *shall* although it is not often used in colloquial Standard English. There are no instances of *shall* in the native speaker data.

- (38) They *shall* say the swiss have good organized hospitals on a high level and they *shall* say, they are (h, German, e-mail)

There did not appear to be any types of non-native tokens which were associated with a particular speaker group, so even in terms of interference-based variation the three groups are very similar. Table 5.8, above, demonstrated that the three groups had similar rates (between 6 and 9%) of interference-based tokens.

## 5.8 Discussion

The above analysis has revealed the following points about the use of *will* and *going to* by the non-native Swiss speakers:

- the rates of *going to* use are far lower in the non-native speakers than in native speakers.
- none of the non-native groups show a change in *going to* use between e-mail and oral data, native speakers show a shift.
- despite there being statistically significant differences between the non-native groups in terms of overall distribution, they are far more similar to each other than they are to the native speakers.
- grammatical person has no effect on the Italian and French speakers, but the German speakers have the native English pattern found by Poplack and Tagliamonte (1999:335), whereby non-first person subjects have more *going to*
- in terms of contraction the non-native groups also show themselves to be considerably different from the native speakers and yet similar to one another.
- considering only *will*, all four groups contract more in first person contexts than elsewhere, non-native overall contraction is lower however.

- interference-based variation tokens are found in all three non-native groups at similar rates; thus the use of future variants does not appear to be influenced by the source languages.

This feature presents some aspects where the non-native speakers match the native speakers, but there are also a number of ways in which the non-native are not following native patterns (see table 5.11). None of the non-native groups have the oral/e-mail shift found in the native speakers and have far lower overall rates of *going to* than native speakers. Register is a very important factor in the use of *going to* in native English, so the overwhelming impression is that the non-native speakers have not acquired the native patterns. This is despite the fact that they pattern similarly to the native speakers in a few ways. This feature cannot be considered one where the native speakers and non-native speakers share the same patterns.

Table 5.11: Summary of similarities and differences in distribution

	Native English	French	German	Italian
Percentage of <i>going to</i>	12 (e-mail) 20-45 (oral)	1	7	2
Oral vs. E-mail distinction	Yes	No	No	No
Grammatical Person	Non-first person favours <i>going to</i>	No pattern	Non-first person favours <i>going to</i>	No pattern
Percentage of <i>will</i> contraction	23	11	12	19
Percentage of <i>going to</i> contraction	8	14	23	29
<i>Will</i> contraction	1 <sup>st</sup> person > other	1 <sup>st</sup> person > other	1 <sup>st</sup> person > other	1 <sup>st</sup> person > other

How do the three non-native groups compare to each other then? This is important as it will allow us to establish whether this feature is involved in focussing or whether the non-native groups share different patterns. As discussed, the three groups have low overall rates of *going to* and do not increase their use of it in speech. All three groups contract *will* less often than *going to* and show a number of similarities in terms of interference-based variation. The German speakers have a grammatical person effect which did not operate for the Italian and French speakers, so there are some differences.

Because a large part of the non-native patterns are found in all three groups, this feature looks as if it is involved in focussing. This conclusion needs to be taken with caution, however, as the low number of *going to* tokens in the three groups mean that the results cannot be easily verified.

Case B	French, German and Italian vary unlike native speakers but like each other	focused form (or common problems of acquisition)
-----------	--	--

## 5.9 Conclusion

In terms of the acquisition of non-native sociolinguistic competence, this chapter has revealed, first of all, that Swiss speakers of English do not make use of the full range of variability found in the future and in many ways do not have one of the variants (*going to*) in their repertoire. Secondly, it has uncovered that the three non-native groups are similar in a number of ways, marking this feature out as a potential case of focussing.

Thirdly, it has shown that although Swiss English classes teach their students some of the differences between the two variants, the non-native groups have not acquired this variability. If they were using the patterns natively then there would have been a higher rate of *going to* in their speech, which there is not.

Moreover, similarities to English in the structure of the future in some of the source languages do not appear to have influenced their acquisition either. This conclusion comes from the fact that the French speakers who have a very similar construction to English *going to* do not use the English variant any more frequently than the other two groups. The speakers of Italian, which does not have any syntactic structures comparable to English, are not very different from the other two groups.

## 6. Relatives

### 6.1 Introduction

English native speaker relative pronoun choice has been the subject of a considerable amount of research (Ball, 1996, Quirk, 1968 and Tagliamonte, Smith and Lawrence, 2005 among others). As stated by Guy and Bayley (1995:148), relative pronoun choice is a ‘clear case of syntactic variation’ in native English which ‘from intuitive reflection [can be said to be] subject to strong constraints’ (Guy and Bayley, 1995:148). As such, it is a useful feature to consider in terms of non-native variation, both because of this previous research but also because the variation has been shown to be conditioned in very specific ways. This analysis will be able to establish the extent to which non-native speakers follow these constraints. Moreover, relative clauses are ‘fairly frequent in speech as well as writing’ (Guy and Bayley, 1995:148), something which is important in the present study as the primary source of data is from a written register.

This variable presents an intriguing site for study as in native speaker data it involves both categorical and variable constraints as demonstrated in (1) and (2).

- (1) the people that/who/\*which are registered (modified from \*, French, e-mail)
- (2) the infos that/which/\*who are in the booklets (modified from b, Italian, e-mail)

Alongside the categorical rules governing relative pronoun choice, there are a number of factors which influence the choice of these pronouns. In terms of non-native teaching of relative clauses, unsurprisingly, only the prescriptive and categorical rules are taught, whilst the variable rules are presented as being *free variation*. By this I mean that the non-native speakers are simply taught that several pronouns are acceptable in the same situation. The aim of the present analysis is then two-fold; it will focus on establishing the extent that the non-native speakers differ from the prescribed norm (so their level of interference-based variation) whilst at the same time judging how closely they follow the variable norms (i.e. target-based variation).

Considering a feature such as this one, where the variability is only partially taught will enable us to determine to what extent non-native speakers are aware of and capable of

dealing with variable rule patterns. Because the full variable rules are not taught, if the non-native speakers are using them, they could have learnt them only from modelling what the native speakers were doing.

The section directly below will present the categorical restrictions. The factors which, instead, only are tied to *preferences* in terms of relative pronoun selection, and as such belong to native speaker variable rules, will be examined in the section 6.6, which also deals with the extraction and coding of the tokens.

## 6.2 Relative Clauses

Relative clauses fulfil similar functions to conjunctions, in that they serve to link two or more clauses together (Quirk et al., 1985:1244ff.). The two clauses in (3) are joined together in (4):

(3) I saw the man. The man was wearing a red hat.

(4) a. I saw the man who was wearing a red hat.

b. The man whom I saw was wearing a red hat.<sup>81</sup>

English has a number of different relative pronouns (*who*, *whom*, *which*, *that*, *when*, *why*, *where*, *whose*), as well as a zero form, which can all be used in various circumstances. Examples 5 to 9 illustrate the use of some of these pronouns using data from the Swiss e-mails.

(5) i've found someone **who** is willing to do the homepage (h, German, e-mail)

(6) I need deans, professors, doctors **to whom** you have a good contact (c, Italian, e-mail)

(7) I think this is a good opportunity to enlarge our activities, especially with the help of an organisation **which** is already settled (b, Italian, e-mail)

(8) Despite the incredible tiredness **that** fell over us in the last days (A, German, e-mail)

(9) a document **Ø** we sent to our Doyen in Geneva (f, French, e-mail)

---

<sup>81</sup> The examples which do not give the speaker code and language were made up for illustration purposes.

The selection of relative pronouns by native speakers is governed by a number of prescriptive rules; for example, although *that* can be used with any type of subject, *which* is restricted to inanimate subjects, while *who* can only be paired with animate subjects (as demonstrated in examples 1 and 2).

It is also important to consider whether a relative clause is **restrictive** or not when considering the selection of pronouns. If the clause which is relativized provides no vital information, then the clause is classed as a non-restrictive relative phrase (example 10), whilst if it does provide necessary information it is classed as restrictive (example 11) (Quirk et al, 1985:1245).

(10) My sister, who is fifteen, went to Rome last year.

(11) The girl whom I was telling you about is his sister.

Convention, in written English, has it that non-restrictive relative clauses are signalled by the use of commas, separating the additional information from the rest of the sentence (Quirk et al, 1985:1245). The distinction between the two type of clauses has been said to be ‘crucial for quantitative studies of relative markers, because NRRs [i.e non-restrictive relatives] strongly favour WH-<sup>82</sup> forms in standard English whereas RRs [i.e. restrictive relatives] do so variably’ (Ball, 1996:229).

In addition to aspects of relative pronoun selection linked to restrictiveness, there are two main properties restricting the section of relative pronouns in English. Relative pronouns can:

- (i) ‘show concord with [their] antecedent, ie the preceding part of the noun phrase of which the relative clause is a postmodifier [external relation]’ (Quirk et al., 1985:1244);

**and**

- (ii) ‘indicate [their] function within the relative clause either as an element of clause structure (S, O, C, A), or as a constituent of an element in the relative clause [internal relation]’ (Quirk et al., 1985:1244).

---

<sup>82</sup> That is to say the forms *who*, *which* and *whom*.



### External relation

The factor which is designated as *external relation* in Quirk et al.'s terminology, is reflected in English relative clauses 'on the basis of a two-term 'gender' system, personal and nonpersonal, and applies only to the *wh*-series' (Quirk et al., 1985:1245). Although *that* and the zero form can both be used either with animate or inanimate objects (examples 12 and 13), the two *wh*- forms are distinguished in that *who* (and its oblique form, *whom*) is used solely with animate subjects or objects, whereas *which* is used only with inanimates (examples 14-17).

- (12) The girl *that/Ø* I saw is pretty
- (13) The toy *that/Ø* I saw is new
- (14) The girl *whom* I saw is pretty
- (15) \*The girl *which* I saw is pretty
- (16) \*The toy *whom* I saw is new
- (17) The toy *which* I saw is new

### Internal relation

In terms of what is designated as *internal relation* by Quirk et al, there are three main types of relative clauses; those with the relative pronoun as subject (example 18), those with it as a direct object (example 19), and those with it as a prepositional object (example 20).

- (18) The man *who* won the prize is very smart
- (19) The man *I* saw is very smart
- (20) The man *to whom* I talked is very smart

*That* and *which* can be used in all three types of relative clause (examples 21-23). *Who* can be used in all three types as well, but only if it is modified to *whom* in object and prepositional relative clauses (examples 24-26). This use of *whom* is rather formal and it is usual, in informal styles, to find *who* used in non-subject relative clauses (example 27). In Standard English, zero, unlike the other relative pronouns, can only be used in non-subject clauses (examples 28-30).

- (21) The toy *that/which* was broken was mine

- (22) The toy that/which I broke was yours
- (23) The box that/which that the toy was in was red
- (24) The man who won the prize is very smart
- (25) The man whom I saw is very smart
- (26) The man to whom I talked is very smart
- (27) The man who I talked to is very smart
- (28) The man I saw is very smart
- (29) The man I talked to is very smart
- (30) \*The man won the prize is very smart

Although all four relative pronouns are acceptable in prepositional relative clauses of certain types, the structure of the sentence is modified in different ways depending on which pronoun is used. Prepositional relative clauses can either be pied-piped or stranded and relative pronouns can be distinguished depending in which of these two structures they are grammatical. In pied-piped relative clauses the preposition is placed immediately before the pronoun (example 31) whereas in stranded relative clauses the preposition is left until after the verb in the subordinate clause (example 32) (Quirk et al. 1985:1246, Tagliamonte, Smith and Lawrence, 2005:95)<sup>83</sup>.

- (31) The man *to whom* I talked was very nice
- (32) The man Ø I talked **to** was very nice

The relative pronoun *that* and the zero form can only occur in prepositional clauses that are of the stranded type, while *which* and *who(m)* can appear with both stranded and pied-piped clauses. Stranded prepositions are frowned upon in writing and in formal speech despite the fact that they are frequently found in everyday conversation (Quirk, 1957:100).

Table 6.1 presents the relative pronouns used in Standard English depending on animacy and the syntax of the embedded sentence. It demonstrates that although the internal and external relations place a number of constraints on the choice of relative clauses, there is nonetheless a considerable amount of leeway, in that there are always at least two possible variants for any given type of relative clause.

---

<sup>83</sup> Studies of spoken data reveal that tokens of pied-piping are very rare; *whom* accounted for a mere 0.05% of the tokens in Tagliamonte, Smith and Lawrence (2005:87).

Table 6.1: Possible relative pronouns					
	Who	Which	That	Zero	Number of pronouns considered grammatical
Subject					
Personal	√	*	√	*	2
Non-personal	*	√	√	*	2
Object					
Personal	√	*	√	√	3
Non-personal	*	√	√	√	3
Prepositional					
Personal	√	*	√	√	3
Non-personal	*	√	√	√	3

These are the prescriptive rules of Standard English for relative pronoun selection and these are the ones which are taught to non-native students of English. Swiss students will have learnt which pronouns not to use depending on the context, but there is no focus on which of the remaining possible forms are most likely to be used by native speakers.

The next section will briefly present a few points about the syntax of relative clauses in the source languages, as this may, at a latter stage, provide some insight into how the non-native speakers acquired and dealt with English variability.

### 6.3 Relative clauses in the source languages

The structure of relative clauses in the three source languages is rather different from that of English, as well as from one another. As there are so many factors at play in the choice of relative pronoun in English, the factors that play a part in relative pronoun selection in the source languages as well need to be noted, in order to gain better insights into their potential interference with English.

The aim of this section is not to fully present the syntax of relative pronouns in the three Swiss languages, but rather to briefly introduce the main aspects of each in order to establish what the similarities or differences in terms of relative pronouns are between each language and English.

#### French

In French, the main relative clauses are *qui*, *que* and *dont*.<sup>84</sup> The choice of pronoun in French is determined almost entirely by the grammatical function of the relative; presenting the matter very schematically, *qui* is used as a subject relative, *que* as an object relative and *dont* as a prepositional object relative. *Qui* can also be used in prepositional relative clauses when in conjunction with a preposition. Animacy does not dictate the form of the relative pronoun selected.<sup>85</sup> In fact, in French, there is no possible variation in relative pronoun choice, as for any given sentence with a relative clause only one variant is acceptable. Furthermore, there is no comparable form to the English zero form in Standard French.

Table 6.2: Examples of French relative pronouns

Subject relative - *QUI*

La fille *qui* est venue (animate subject relative – feminine subject)

Le courrier *qui* est arrivé (inanimate subject relative – masculine subject)

Object relative – *QUE*

La fille *que* j’ai vue (animate object relative – feminine object)

Le courrier *que* j’ai reçu (inanimate object relative – masculine object)

Prepositional relative - *DONT*

La fille *dont* je t’ai parlé (animate prepositional relative – feminine object)

Le livre *dont* je t’ai parlé (inanimate prepositional relative – masculine object)

Prepositional relative – *QUI* + preposition

La fille à qui j’ai donné une rose (animate prepositional relative – feminine object)

Le courrier auquel j’ai répondu (inanimate prepositional relative – masculine object)

## Italian

Generally speaking, there is one main relative pronoun in Italian, *che*. This pronoun is used for animate and inanimate objects and for direct and indirect objects. For prepositional clauses, a determiner marking the gender of the relative subject with *quale* can also be used. As in the case of French, Italian has no zero form of the relative pronoun.

<sup>84</sup> Like English, there are also a number of adverbial relative clauses, (*où*, *quand*, *etc.*) but as it was decided to leave them aside in English, I will not deal with the French ones (or those of German and Italian) either.

<sup>85</sup> Note that it does however affect the verb form in some tenses as in the examples in table 6.2.

Table 6.3: Examples of Italian relative pronouns

Subject relative - CHE

la ragazza *che* è venuta

il libro *che* è caduto

Object relative - CHE

La ragazza *che* ho vista

Il libro *che* ho comprato

Prepositional relative – Det + QUALE, CHE, Prep + cui

La ragazza *la quale/che* mi piace

Il fiore *il quale/che* mi piace

La ragazza *a cui* ho dato il libro.

### German

German, like French, makes a difference between direct and indirect objects and does not distinguish animate from inanimate. The case and gender of the relativized noun is important, however, because the relative pronouns are in fact determiners.

In Swiss German, the main form of relative pronoun is *wo*, and this is used for all noun cases. This is used alongside the German forms.

Table 6.4: Examples of relative pronouns in (Swiss) German

Subject relatives – DER, DIE, DAS

Der Mann, *der* gekommen ist

Die Frau, *die* gekommen ist

Das Mädchen, *das* gekommen ist

Object relatives – DER, DIE, DAS (+declension)

Der Mann, *den* ich gesehen habe

Die Frau, *die* ich gesehen habe  
 Das Mädchen, *das* ich gesehen habe

WO

dr Maa wo mer geschter gseh hei (Penner and Bader, 1995:135)

dr Maa wo mer geschter mit-em gschpiut hei (Penner and Bader, 1995:135)<sup>86</sup>

Overall, the three Swiss languages have different strategies in terms of relative pronoun use and these strategies are for the most part different from English relative pronoun strategies.

## 6.4 Previous Studies

Quirk et al. (1985) provides a useful starting point for the discussion of factors influencing the selection of relative pronouns and deals with a number of factors which will be considered in the subsequent analysis. However, as its focus is mainly typological and its discussion of variability based on intuition rather than empirical study,<sup>87</sup> we must also consider the work of other researchers who have looked into relative pronoun selection.

There is a considerable body of literature on relative pronouns studied using variationist methods. These studies consider relative pronouns in terms of native speakers from a number of different regions, including American English (Biesenbach-Lucas, 1987, Kikai, Schleppegrell and Tagliamonte, 1987, Guy and Bayley, 1995, Ball 1996), early African American English (Tottie and Harvie, 2000, Tottie and Rey, 1997), educated British English (Quirk, 1965), ‘Dorset’ English (Van den Eynden, 1993), varieties of Northern British English (Tagliamonte, Smith and Lawrence, 2005), non-native English (Flanagan and Inal, 1996, Karlstadt, 1996) and second generation Mexicano-English (Bayley, 1999).

The studies mentioned above are not all equally valuable to the present inquiry, as the methodology of some means that they are not adequately comparable to the present data. Biesenbach-Lucas (1987), for example, does not distinguish relative pronouns by syntactic type. Thus, her results are not comparable to any of the other studies, because syntactic

<sup>86</sup> Translation of the Swiss German sentences ‘the man that we saw yesterday’, ‘the man whom we played with yesterday’.

<sup>87</sup> Although it is reasonable to assume that Quirk’s own empirical research (1965) into relative pronouns may well have been the basis of a number of the comments in the later work.

type restricts the pronouns available and unless disentangled the distribution of the pronouns is obscured. Similarly, Flanagan and Inal's research into non-native relative pronouns was conducted through questionnaires asking which relative pronouns the informants would use and thus does not permit us to uncover the variable rules in their informants' speech. For this reason, only the studies which have analysed their data in a manner that allows us to establish the distribution of forms according to the various constraints on use in terms of restrictiveness, syntax type and animacy will be considered.

An analysis of the findings of previous studies has been particularly useful in establishing which factors have been shown to influence (or potentially influence) the choice of relative pronouns. All the factors examined in the present analysis have been found to be significant in other research, which means we can verify the extent to which the non-native speakers are comparable to native speakers. Because the findings of these previous studies were crucial to the selection of factor groups in this analysis, they will be presented in detail in section 6.6 when discussing the factors considered in this study.

## 6.5 Data Sources

The non-native tokens for this feature come from the e-mail corpus, so the native e-mail corpus is the main point of comparison. Although the tokens from the IFMSA oral corpus were extracted as well, there were not enough of them to enable us to compare the two registers adequately and thus they will not be considered in this analysis.

## 6.6 Extraction and coding

This section will focus on two aspects; first of all, it will discuss circumscribing the variable context, and secondly, this section will present the factor groups which were selected for analysis.

### Exclusions

Because the aim in the present chapter is to consider variability in relative pronoun selection, only the relative pronouns which can (under the right circumstances) be substituted for one another were considered. The pronouns which fulfil this requirement are *who*, (*whom*), *which*, *that* and the zero form. Adverbial relative pronouns such as *where*,

*when*, *why* have been excluded from consideration as they cannot be as readily exchanged with the other variants.<sup>88</sup>

Furthermore, because non-restrictive relative clauses do not allow for the same range of pronouns, they too have been excluded from the present analysis.<sup>89</sup> Both non-restrictive relative clauses and adverbial relative pronouns were excluded from analysis in a number of previous studies (Guy and Bayley, 1995:150, Tagliamonte, Smith and Lawrence, 2005:85 for example) so this will make it easier to compare the present results with them.

As the main focus of the analysis of this feature rests in the cases where a number of variants are possible, those cases where variability was constrained were removed. As such only adnominal relative clauses (33), considered the ‘central type of relative clause’ in Quirk et al (1985:1244) were selected, leaving nominal (34) and sentential (35) relative clauses aside.

(33) The news *which* appeared in the papers this morning was well received.

(34) *What* surprises me is that they are fond of snakes and lizards.

(35) They are fond of snakes and lizards, *which* surprises me.

(all three examples from Quirk et al., 1985:1245)

### Factor Groups

The relative tokens extracted were coded for a range of factor groups; first of all, a number of **external** factors were considered:

- *e-mailer*

- *e-mailer's native language*

- *register (oral vs. e-mail)*

<sup>88</sup> *Whose* has also been excluded because it cannot be substituted by the other relative pronouns. However, it did appear to present some difficulty for the non-native speakers and will be touched on briefly in the analysis section of this chapter.

<sup>89</sup> Although Standard English constrains the use of *that* with non-restrictive relative clauses there are some cases of this in the non-native data:

- as well as explaining to ‘f’ or ‘g’, **that** will be in Malta, what to do for you during the SCORP Working Committee (b, Italian, e-mail).

This token and ones similar to it were excluded from the analysis as they are nonetheless non-restrictive.



This last factor was important to consider because previous studies have shown that the selection of relative pronouns is influenced by formality. Quirk et al. state that ‘all things being equal, more informal discourse will tend to have a preference for zero’ (1985:1252). Moreover, the *wh*- forms are more commonly used in writing than orally (for example, Tagliamonte, Smith and Lawrence, 2005:87 find that the *wh*- forms are virtually inexistent in their corpora of vernacular dialects). Until we have examined the e-mail data, we cannot establish whether in this instance it is more similar to speech or written data.

The other external factors have already been discussed in the coding section of other features (especially chapter 4), so will not be dealt with further at this time. Alongside these external factors, a number of **internal** factor groups were considered as well.

The factors discussed in the introduction to relative clauses, i.e. those place prescriptive rules on pronoun selection, will need to be considered by examining the results of each category separately. The main factor groups which restrict the pool of potential relative pronouns are:

- (1) animacy
- (2) the function of the relative pronoun<sup>90</sup> (Quirk et al., 1985:1248)

As well as prescriptively restricting the pool of acceptable pronouns, these two factors to variably constrain relative pronoun choice.

### ***Animacy***

Animacy was chosen as a factor, because humanness or non-humanness of the pronoun being relativized has been found to play a crucial role in relative pronoun selection in English. As discussed in section 6.1, it helps determine which pronouns are available to use, but even in terms of the allowed pronouns, other research (Guy and Bayley, 1995, Quirk, 1964 and others) has found that this is variable. For example, previous research has found that animate subjects are more likely to occur with the *wh*- form than with *that* or zero (Tagliamonte, Smith and Lawrence, 2005:17).

---

<sup>90</sup> Restrictiveness was also considered in Quirk et al. as a factor constraining relative pronoun selection; however, as discussed earlier non-restrictive relative pronouns have been excluded from the present analysis for the reasons noted.

In terms of the present analysis, individuals or groups of people (example 36) were considered to be human and everything else to be non-human<sup>91</sup>.

- (36) In particular the people *who* will come to Denmark (b, Italian, e-mail)

### ***Syntactic function***

The second factor considers the type of relative clause; distinguishing subject relatives from object relative and prepositional relatives. This factor group is divided into two factors: subject relatives and object relatives.

Unlike in a number of other studies, the various types of object relatives were not considered as separate categories within this factor group because the main distinction in terms of syntax and relative pronoun selection is a two-way division. Moreover, the data sample overwhelmingly contained subject relatives. Object relative types were considered, however, in an additional factor group. This allowed me to examine whether non-native speakers had similar patterns to native speakers in terms of object relatives.

Syntactic function constrains pronoun choice as follows ‘when the antecedent is personal and the pronoun is the subject of the relative clause, *who* is favoured, irrespective of style and the occasion’ (Quirk et al., 1985:1250).

### ***Type of object relative***

This factor group distinguished the type of object relative: direct object relatives (example 37), prepositional object relatives with stranding (38) and prepositional object relatives showing a pied-piped construction (39).

- (37) I won a flight *which* I cannot refuse! (d, German, e-mail)  
 (38) write a kind letter to one of the profs Ø you are in good contacts with at the hospital (b, Italian, e-mail)  
 (39) I need deans, professors, doctors *to whom* you have a good contact (c, Italian, e-mail)

Stranded and pied-piped relatives do not allow the same relative pronouns. Furthermore, as noted by Quirk et al. (1985:1251), when the antecedent is personal but the object of a verb

---

<sup>91</sup> There were no tokens of animate but non-human subjects (i.e. animals or aliens) so this was not given a factor code.

or prepositional complement ‘there is a much stronger preference for *that* or zero, perhaps to avoid the choice between *who* and *whom*’.<sup>92</sup>

Two further factors were tested in this study, based on the findings of previous studies and a preliminary analysis of the non-native data.

### ***Definiteness***

This factor focuses on the definiteness of the antecedent NP head, whether it is a definite noun phrase (example 40) or an indefinite noun phrase (example 41).

(40) here is the address of the man *who* was in charge of organising (b, Italian, e-mail)

(41) How we must speak with someone *who* wants to stop smoking (f, French, e-mail)

Van den Eynden notes that ‘the restrictive *wh*- and *th*- forms occurred most frequently with indefinite antecedents’ and that ‘zero, on the other hand, occurred exclusively in definite contexts’ (1993:113).<sup>93</sup>

The coding followed Van den Eynden’s division; whereby ‘nouns modified by the definite article, nouns preceded by a demonstrative, possessive, or interrogative (what/which) determiner, proper names, and personal, demonstrative, possessive and interrogative pronouns’ (Van den Eynden, 1993:116) were considered to be definite, while ‘nouns modified by the indefinite article, the zero article, by a cardinal number, or by a quantitative determiner, and indefinite or quantifier pronouns’ (Van den Eynden, 1993:116) were considered indefinite.

### ***Adjacency***

---

<sup>92</sup> Quirk and al. adds ‘*whom* would seem pedantic to many people, while *who* as object in relative clauses is informal and tends to be regarded as incorrect. Since, therefore, neither *who* nor *whom* is wholly satisfactory, *that* (and particularly zero) is frequently used despite a personal antecedent’ (1985:1251). Moreover, ‘avoidance of *whom* may not be the only factor influencing *that* as object with personal antecedent. Grammatical object are more likely to be nonpersonal, or to carry nonpersonal implication, than subjects’ (1985:1252).

<sup>93</sup> Although the original coding considered pronouns as a separate factor within this group (as had been done in Bayley, 1999), there were too few tokens of pronouns in the data so non-native tokens were recoded as being either definite or indefinite.

A number of studies found adjacency to be important for the selection of relative pronouns. The zero form has been found to be more likely to occur in adjacent position than in other positions. Indeed, Quirk et al. (1985:1252) have stated that ‘when complex phrases or clauses intervene between the antecedent head and the relative pronoun, *which* is generally preferable to *that* and very much preferable to zero’. Note that because Quirk et al. were considering both restrictive and non-restrictive pronouns, this difference may be related to that in part. Guy and Bayley (1995) found this factor to be significant in their analysis, however, and they had excluded non-restrictive relative pronouns.

This factor was divided into three groups, following Guy and Bayley (1995); relative pronouns adjacent to their antecedents (example 42), relative pronouns separated from their antecedents by another relative clause (example 43) and relative pronouns separated from their antecedents by a phrase other than a relative clause (example 44).

- (42) If anybody knows somebody *who* is keen in design. (b, Italian, e-mail)
- (43) for those who are arriving on the 3. or 4. march in malta (or *who* are staying longer) (h, German, e-mail)
- (44) She also adviced to have a nurse or somebody working with older people *that* can bring us (b, Italian, e-mail)

Quirk et al. (1985:1248) discussed a number of other factors which could influence the choice of pronoun, such as the ‘medial or final position of the relative clause in relation to the superordinate clause, and the length of the relative clause’ but as these were not found to be significant in the studies using multivariate analysis (see for example, Bayley, 1999 and Tagliamonte, Smith and Lawrence, 2005), they have not been considered in the present study.

Table 6.5 on the following page presents an overview of the factors considered for analysis in a number of earlier studies which examined relative pronouns.

<b>Table 6.5</b>		Quirk	Biesenbach- Lucas	Kikai, Schlepppegrell, Tagliamonte	Van den Eyden	Guy and Bayley	Ball	Flanigan - Inal	Karstadt	Tottie and Rey	Bayley	Tagliamonte, Smith and Lawrence
(factors included)		1957	1987	1987	1993	1995	1996	1996	1996	1997	1999	2005
Multivariate Analysis				x		x		x	x	x	x	x
Syntactic Function of the relative pronoun in the relative clause	subject, object, pied-piped...	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Humanness/Animacy		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adjacency		✓			✓	✓	✓			✓	✓	(✓)
Category of Subject of relative clause	pronoun, noun, relative pronoun	✓								✓	✓	
grammatical category of antecedent NP head	definite, indefinite, pronoun	✓		✓	✓		✓			✓	(✓) <sup>94</sup>	✓
Special lexical cases - indefinite/definite					✓		✓					
Premodification	no premodifier, premodifier, nonpremodifiable									✓	(✓)	
Syntactic Function of the NP head in the matrix clause	subject, direct object, object of preposition...					(✓)				✓	(✓)	
Length		✓			✓		✓					✓
Position in the higher clause		✓			✓		✓					
Restrictiveness			✓	NR excluded	✓		NR excluded	NR excluded	✓		NR excluded	NR excluded
Channel of Communication			✓			✓		✓				
Non-native vs. native								✓				

<sup>94</sup> Bracketed factors are those which were considered but found not to be significant in the multivariate analysis.

## 6.7 Results and Analysis

The extraction of the native and non-native e-mail corpora yielded over 800 tokens. As can be seen in table 6.6 below, each group has at least 100 tokens of various relative pronouns<sup>95</sup>.

Table 6.6: Number of relative pronoun tokens by native language	
English	266
French	149
German	109
Italian	289
Total	813

The overall number of tokens per group does not show us much in terms of how relative pronouns are selected. Table 6.7 below provides the overall distribution of the four most frequent relative pronouns in the data by language group. The fifth relative pronoun considered in this analysis, *whom*, did not appear in all the language groups and was used rather infrequently so is not presented here; however, it will be dealt with in the discussion of object relatives.<sup>96</sup> Figure 6.1 presents the same data in graph form to make it more readily interpretable.<sup>97</sup>

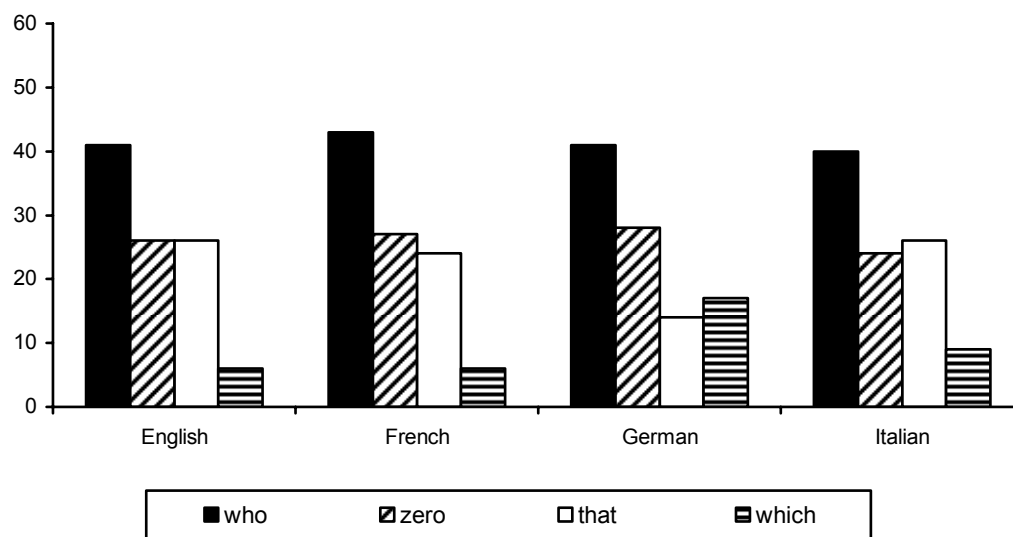
Table 6.7: Overall distribution of the most frequent relative pronouns in data by language group (%)					
	<i>who</i>	<i>that</i>	<i>which</i>	<i>zero</i>	Total N
English	41	26	6	26	266
French	43	24	6	27	149
German	41	14	17	28	109
Italian	40	29	9	24	289

<sup>95</sup> The oral corpus yielded 18 tokens for the German speakers, 23 for the Italian speakers and 81 for the French speakers.

<sup>96</sup> This is why the numbers in tables 6.6 and 6.7 do not match for the French and Italian speakers.

<sup>97</sup> A graph will be provided alongside all the subsequent tables as well.

### 6.1: All relatives



All four groups use *who* most frequently and, except for the German speakers, *which* least frequently; so they look the same in terms of frequency of use. Remember, however, that the distribution of relative pronouns is highly constrained in terms of clause type and in terms of the animacy of what is relativized.

It is only when these effects have been disentangled that we can determine how similar the native and non-native groups really are. Before considering this however, we must focus on the interference-based aspects of relative pronoun selection. If the results reveal that a large proportion of the data in the analysis is different from Standard English, then this will significantly affect our results as well.

#### Interference-based variation

The analysis revealed that there were in fact very few cases of interference-based variation in the Swiss data; only 11 in out of a total of 547 tokens (2% overall). Six tokens were found where *which* was used with animate subjects (all in the output of Italian speakers) (examples 45 and 46), while five tokens of *who* were used with inanimate subjects (three in a French speaker, two in an Italian speaker) (example 47).

- (45) A NEO is not only a responsible person, but also one **which** regularly gives time to the associations, **that** knows (c, Italian, e-mail)
- (46) There where a lot of new Members, especially from Asia, Eastern Europe and South America, **which** had to be introduced into the(c, Italian, e-mail)
- (47) Ex. I forgot to mention that the magical letter **who** went directly into the hearts of the director ... (b, Italian, e-mail).

These non-native tokens represent only 2% of the overall tokens of relative pronouns in the data, which may explain why it is not a feature considered by Second Language Acquisition researchers. From a purely prescriptive perspective, the non-native speakers have acquired the feature natively.

The initial analysis found another relative pronoun that created a considerable amount of difficulty for the non-native speakers. *Whose* was not selected as one of the relative pronouns under study because it did not occur frequently enough in the data and also because it was not fully interchangeable with the other variants. There are, however, a number of points to be made in terms of its non-native use. In fact, it is more a case of non-use in terms of the non-native speakers, as the extraction did not yield a single token of *whose* for them, whilst 13 tokens of it were found in the native speakers. Thus, we cannot establish if it is a case of the non-native speakers avoiding the use of a form they felt uncomfortable with or whether more data, or different data, would have produced more tokens of it. A few of the sentences discarded from the analysis, however, showed cases of sentences where a native speaker would have used *whose* but where the non-native speakers used another strategy (examples 48-49) so it looks as if they did not know how to use it in a native English manner.

- (48) The only way to look for a computer **which** address is not indexed / known (such as yours) is, when you are (j, French, e-mail)
- (49) the person **where** he live phone me to say she doesn't want to lodge him (f, French, e-mail).<sup>98</sup>

---

<sup>98</sup> It is difficult in this example to ascertain precisely what was meant; however, a native speaker could possibly have used a construction such as 'the person whose house he lives in'. These tokens were of course not included in the analyses presented below.



Overall, the tokens of relative pronouns which are ungrammatical to native English speakers represent a tiny fraction (2%) of the tokens of relative pronouns in the data; so generally, the pronoun choice of non-native speakers is not noteworthy to native speakers. It is only a more in-depth analysis which will be able to tell us how similar or how different the native and non-native groups are. The factors found to affect native variability will be considered in turn below, considering first the prescriptive rules which dictate which pronoun is acceptable in a given circumstance (i.e. the factor groups dealing with animacy and relative clause syntax) and then the variable rules which have been found to influence relative pronoun choice (adjacency, definiteness). The few non-native tokens have been left in the analysis as this may allow us to uncover what factors triggered their use and the low number of them will not affect the results to a great extent.

Animacy will be considered within the analysis of subject and object relatives as they are interlinked in terms of the range of relative pronouns acceptable. Subject relatives allow only *who* and *that* for animates and *which* and *that* for inanimate, while object relatives allow *who(m)*, *that* and zero for animate and *which*, *that* and zero for inanimate. For both subject and object relatives, all tokens of each type will first be considered together, and then animate and inanimate subject/object relative are considered separately. This will allow us to establish in what ways the linguistic groups are similar.

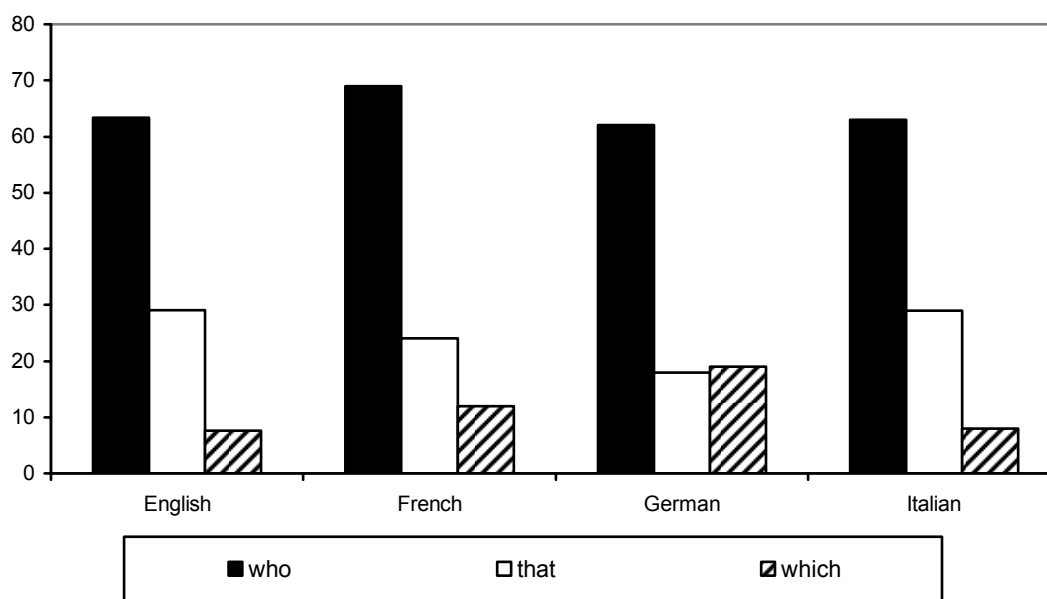
### Subject relatives

Subject relatives make up 63% of the overall tokens of relative pronouns in the data (513 out of 813). Table 6.8 and graph 6.2 provide the overall distribution of subject relatives by speaker groups.

Table 6.8: Overall distribution of Subject relatives in data				
	<i>who</i>	<i>that</i>	<i>which</i>	Total
	%	%	%	N
English	63	29	8	172
French	69	24	7	91

German	62	18	19	72
Italian	63	24	12	178

## 6.2 Subject Relatives only

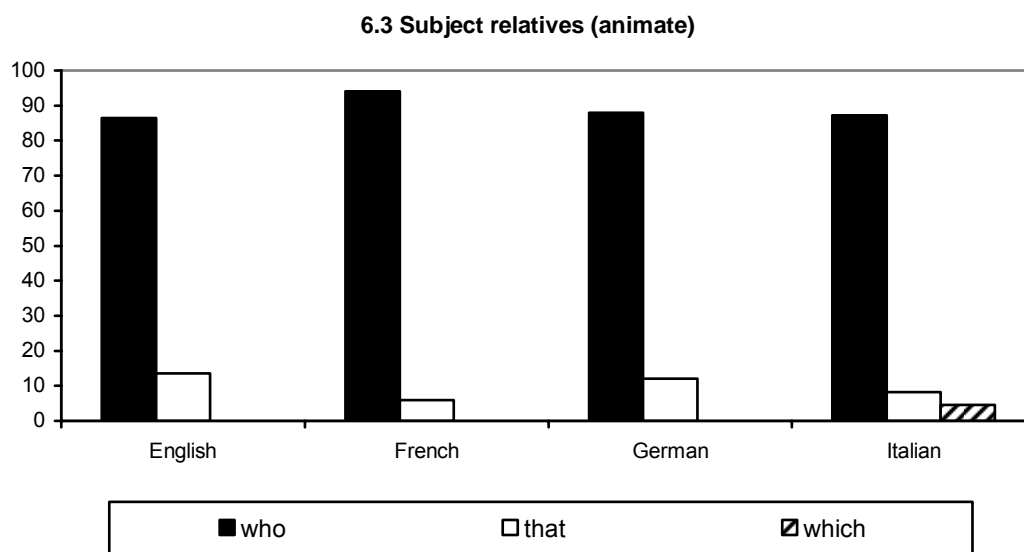


The French, Italian and English groups share the same hierarchy (who>that>which), while the German group use *which* slightly more often than *that*. The German speakers have *who* as their main form, as do the other three groups. Because the relative pronouns used in subject relative clauses are closely linked to the animacy of the subject, human and non-human subjects have to be disentangled before assessing how similar the groups are.

The following tables and charts consider this, looking first at human subjects (table 6.9 and graph 6.3) and then at non-human subjects (table 6.10 and graph 6.4) to determine whether the surface differences in distribution between the groups are reflected in the more detailed analysis. Differences in these tables will be a far more accurate way of judging whether the native and non-native groups have substantially different processes

to deal with the relative pronoun choice than the first tables (such as 6.6) presented in the analysis.

Table 6.9 Overall distribution of Subject relatives in data (human only)				
	<i>who</i>	<i>that</i>	<i>which</i>	Total
	%	%	%	N
English	87	13	-	126
French	94	6	-	63
German	88	12	-	51
Italian	87	8	5	129

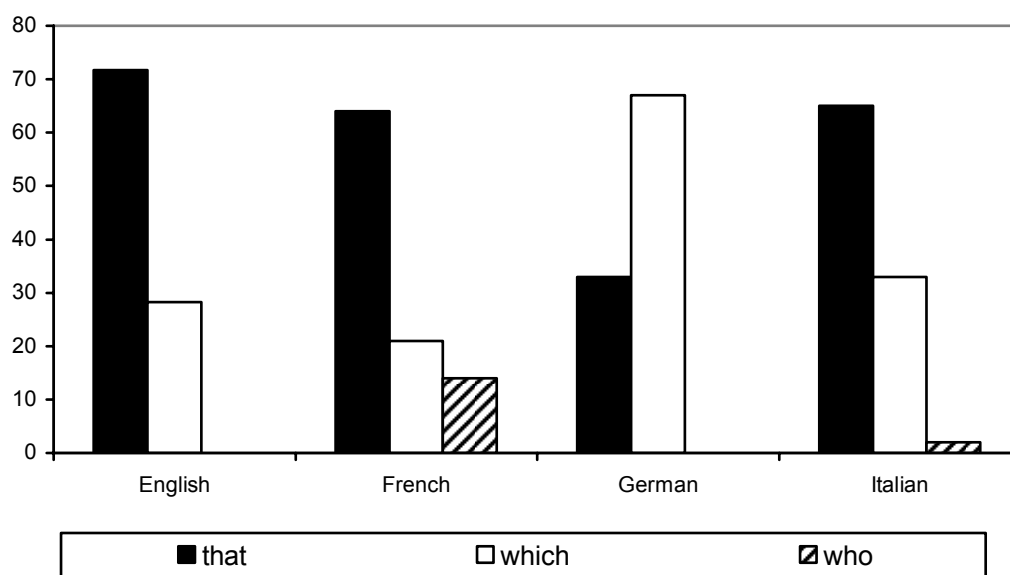


Except for the tokens of *which* found only in the Italian speakers, and which are due to interference-based variation, the patterns of the four groups are quite similar; in all four groups *who* is the main relative pronoun used for human subject relative clauses with *that* being used less frequently.

Table 6.10 Overall distribution of Subject relatives in data (non-human only)				
	<i>that</i>	<i>which</i>	<i>who</i>	Total
	%	%	%	N

English	72	28	-	46
French	64	21	14	28
German	33	67	-	21
Italian	65	33	2	49

#### 6.4 Subject relatives (inanimate)



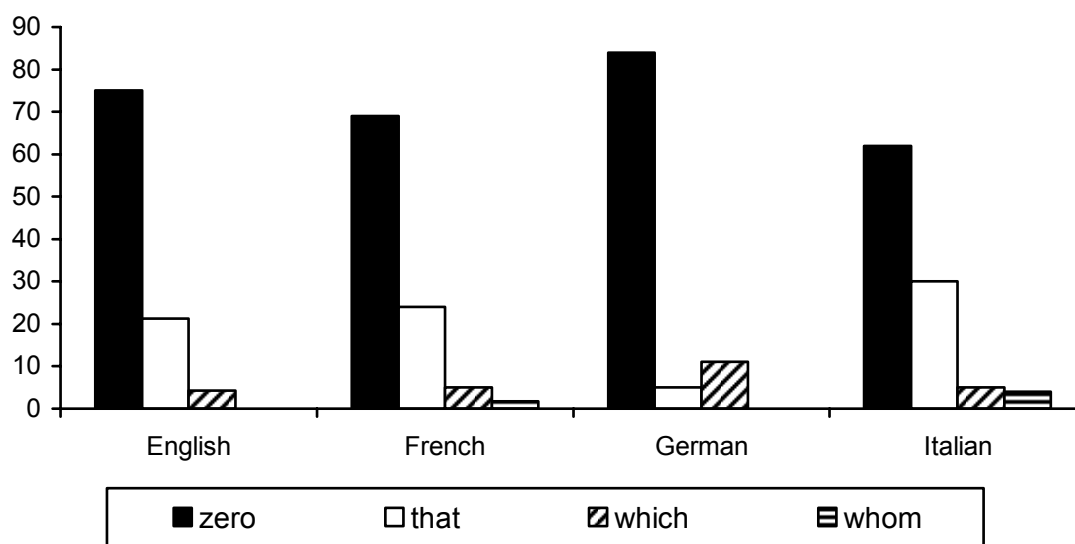
As with the human subject relatives, there are a few tokens of non-native forms (*who*) used by the French and the Italian speakers, but what is most striking about these figures is the difference between the German speakers and the other groups. Whilst the native speakers and the other non-native groups use *that* as the main pronoun for non-human subjects, the German speakers are far more likely to use *which*. Indeed, in terms of actual percentages, the tendencies are reversed, in that the native English speakers use *that* around 70% of the time and *which* the remaining 30% and the German speakers use *which* 70% and *that* 30%. Note, however, that it could in part be due in part to the lower number of tokens found for the German speakers (21 vs. 28 and 49 for the other two non-native groups), but it is nonetheless an intriguing finding.

#### Object relatives

Turning to the object relatives extracted in the data, note that they occurred less frequently than subject relative clauses; making up 37 percent of the total number of relative pronouns 300 out of 813. This is fairly typical; Tagliamonte, Smith and Lawrence (2005:89) found 33% object relative clauses in their data. Table 6.11 and Chart 6.5 demonstrate that there are fewer differences between language varieties in terms of hierarchies for the object relatives than the subject relatives.

Table 6.11: Overall distribution of object relatives					
	<i>zero</i>	<i>that</i>	<i>which</i>	<i>whom</i>	Total
	%	%	%	%	N
English	75	21	4	0	94
French	69	24	5	2	58
German	84	5	11	0	37
Italian	61	30	5	4	111

**6.5 Object relatives**



The English, French and Italian groups all show the same hierarchies of zero > that > which > (whom) and while the German speakers reverse *that* and *which*, the percentage difference between the two is not as considerable as in the case of the inanimate subject relatives and they, like the other three groups, have zero as the main form.

In terms of percentages, all four groups use the zero form over sixty percent of the time and, except for the German speakers, *that* between twenty and thirty percent. The low use of *that* for the German speakers is similar to the findings for the subject relatives, this shows that the German speakers are avoiding its use and are instead going for the next most frequent option instead. In the case of object relatives, the zero form is used overwhelmingly by the German speakers (84%); far more than by the other language groups.

Another aspect which is worth noting is the use of *whom* by the non-native speakers; the native tokens revealed no cases of this oblique form, whilst the non-native data provided one token for the French speakers and three for the Italian speakers. To some extent, this demonstrates that the non-native e-mails are closer to Standard written English, but overall the number of tokens for human object relatives is extremely low (around 20 tokens across the four groups) and this is a factor as well.

Because there are only 21 tokens across the four linguistic groups of human object relatives, the results for the two types separately will not be presented as individual patterns of distribution could not be determined.

What have the results thus far shown us? Although the variable rules governing relative pronoun selection were not explicitly taught to non-native speakers, they nevertheless are adhering to them to a great extent. The following results underline this:

- for **human subject relatives** all four linguistic groups share the same *who*>*that* pattern
- in terms of **non-human subject relatives** the English, French and Italian groups share a *that*>*which* pattern

- in terms of object relatives all the groups have zero as the main variant, followed by *that* then *which* for the English, French and Italian groups and *which* then *that* for the German group.

These results demonstrate that the non-native groups are very similar to the native group and most importantly share the hierarchies and percentages of the native control group. This is very different from the results found for the future (chapter 5), where *going to* was discarded and *will* was essentially the only form in use for all three non-native groups. While at this stage, we cannot determine whether the variable rules are exactly the same, this has shown that the patterning is in the same direction as native speakers.

For the most part, the non-native speakers use the relative pronoun variants in a similar way to native speakers. There are of course a few deviations from the native norm, with unsuitable clauses used for animate or inanimate objects but these represent a very small and unimportant percentage of the overall number of tokens. Although the German speakers use some of the forms (*which*) differently from the other two Swiss groups and the native group, by and large the main form is the same everywhere, furthermore the main divisions, object vs. subject relatives are the same.

The differences found in the German speakers are difficult to account for at this stage; they could be due to interference with their mother tongue or they could be linked to the factors we have not examined yet. I will come back to this in the discussion section.

How do these results compare to what previous studies have found? Ball (1996) provides a summary table of the results of subject relative pronoun choice for a number of earlier studies of Standard English (British and American) divided into human and non-human subjects. Comparing the results from the e-mail data collected for this thesis to the results from other studies will allow for a greater measure of verifiability to help establish the extent to which the non-native speakers do indeed share the native patterns.

As discussed before, although the native e-mails were collected to form a control group, as a counterpart to the non-native e-mail corpora, non-native speakers would have had the most contact with Standard English orally or through formal writing. It may turn out that the few differences which were found between the native and non-native groups are linked to the fact that the native corpus is comprised of e-mails and that the non-native speaker results are in fact closer to native speech.

Table 6.12 presents the results for human subject relatives, while Table 6.13 provides those for non-human subject relatives, considering both American data from the nineteen-seventies and British data from the nineteen-fifties. All four studies discussed by Ball come from Standard, middle class varieties of English, most likely closely approximating what the Swiss students would have been exposed to during their acquisition of English.

Table 6.12: Human subject relative pronouns: comparison of Ball 1996 and present data

			who	that	zero	other	Total
			%	%	%	%	N
Watergate Hearings	70s	Ball (1996)	<b>98</b>	1	-	-	50
Terkel MC interviews	70s	Ball (1996)	<b>92</b>	8	-	-	50
Nixon	70s	Guy and Bayley (1995)	<b>81</b>	19	-	-	114
Standard British English	50s	Quirk (1957)	<b>91</b>	9	0.45	-	222
British Native E-mail	2000		<b>87</b>	13	-	-	126
Non-natives	2000						
French	2000		<b>94</b>	6	-	-	58
German	2000		<b>88</b>	12	-	-	51
Italian	2000		<b>87</b>	8	-	5 (which)	129

Table 6.13: Non-human subject relative pronouns: comparison of Ball 1996 and present data

			which	that	zero	other	Total
			%	%	%	%	N
Watergate Hearings	70s	Ball (1996)	24	<b>76</b>	-	-	50
Terkel MC interviews	70s	Ball (1996)	12	<b>84</b>	-	-	50
Nixon	70s	Guy and Bayley (1995)	13	<b>87</b>	-	-	130
Standard British English	50s	Quirk (1957)	48	<b>52</b>	0.33	-	304
Native E-mail	2000		28	<b>72</b>	-	-	46



Non-natives	2000						
French	2000		21	64	-	14 (who)	28
German	2000		67	33	-	-	21
Italian	2000		33	65	-	2 (who)	49

In terms of human subject relative pronouns, there are no major differences between the e-mail and oral results; in all cases, *who* is the variant used most frequently. In terms of the non-human subject relatives, the hierarchy is the same for all the groups but the German speakers; *that* is the primary variant with *which* as the alternate, but there are greater divisions in terms of percentages than for human subject relatives. Leaving the German speakers aside, *that* is selected between 52 and 87% depending on the group under study. In terms of the native speaker groups, it could be seen to represent a change in progress as the results span nearly fifty years. Quirk's results show the highest use of *which* (48%), while the native British e-mails collected in 2000 and later show only a 28% use of *which*. Backing this hypothesis up further, Tagliamonte, Smith and Lawrence's 2005 research found that the *wh*- forms to be almost totally absent from their data. The non-native speakers are approximating current relative pronoun patterns, not those of fifty years ago. The native results also demonstrate differences between American and British English patterns as the three American studies show more *that* use than the later British e-mail study. The French and Italian distribution is very similar to the distribution in the native groups.

Having compared the native and non-native groups' distributions in terms of animacy and relative clause type, we will now focus the factors of adjacency and definiteness to establish whether the native patterns are replicated by the non-native speakers here as well.

Note, however, that object and subject relatives are considered together for these two factor groups and this may partly obscure what is occurring. They must be considered together at this stage to avoid getting extremely low token numbers in some of the cells, which would make it difficult to establish any patterns. The multivariate analysis will be able to determine the relative strength of the different factors, however.

## Adjacency

The four tables below (6.14-6.17) present the distribution of the three variants in terms of whether the relative pronoun is adjacent to the subject or separated by another relative clause or separated by something else. Each table considers a linguistic group separately. Previous studies have found that the zero form is far more likely to be found in an adjacent position than elsewhere.

The results for both adjacency and definiteness make a three-way distinction of the relative pronouns, combining *who* and *which* into a single form. This is to partly reduce the animacy distinction and also because it is how a number of earlier studies examined relative pronouns and as such this allows for an easier comparison.

Table 6.14: French relative pronoun distribution by adjacency

French	% of wh-	% of that	% of zero	Total N
Adjacent	46	23	31	128
Sep. relative	50	50	0	2
Sep. other	68	32	0	19

6.15: Italian relative pronoun distribution by adjacency

Italian	% of wh-	% of that	% of zero	Total N
Adjacent	51	21	28	237
Sep. relative	57	43	0	14
Sep. other	39	55	5	38

6.16: German relative pronoun distribution by adjacency

German	% of wh-	% of that	% of zero	Total N
Adjacent	55	14	30	92
Sep. relative	100	0	0	2
Sep. other	67	13	20	15

6.17: English relative pronoun distribution by adjacency

English	% of wh-	% of that	% of zero	Total N
Adjacent	46	26	28	239
Sep. relative	25	50	25	4
Sep. other	61	26	13	23

All four tables reveal, first of all, that zero relative pronouns are found more frequently in an adjacent position than when separated from the main clause. Indeed, there are no tokens of zero in non-adjacent position whatsoever for the French speakers. Although the English, Italian and German speakers have some of zero in these positions, it is still lower than for adjacent tokens. The low number of tokens separated by another relative clause in all four groups means that this factor will need to be collapsed with the tokens separated by other in the multivariate analysis. The overall proportion of non-adjacent tokens is rather low and ranges from 10% for the English group to 18% for the Italian group; nevertheless it is high enough to allow us to consider the effect of adjacency on the distribution of the relative pronoun variants.

These results match what had been predicted by previous studies; i.e. the zero form is most likely in adjacent positions. The non-native speakers have acquired this constraint.

### Definiteness

The four tables below (6.18-6.21) present the three variants in terms of definiteness; unlike adjacency, the distribution of tokens between definite and indefinite relative pronouns is better balanced. This means that any conclusions that can be drawn from the distributions will be far more accurate. By comparing the results of definite and indefinite contexts separately for the four linguistic groups, we will be able to establish whether the patterns are similar.

Table 6.18: French relative pronoun distribution by definiteness

	% of wh-	% of that	% of zero	Total N
Definite	37	21	42	71
Indefinite	60	27	13	83

Table 6.19: Italian relative pronoun distribution by definiteness

	% of wh-	% of that	% of zero	Total N
Definite	41	27	32	193
Indefinite	68	25	7	96

Table 6.20: German relative pronoun distribution by definiteness

	% of <i>wh</i> -	% of <i>that</i>	% of zero	Total N
Definite	49	9	42	65
Indefinite	70	20	9	44

Table 6.21: English relative pronoun distribution by definiteness

	<i>Wh</i> -	<i>That</i>	Zero	Total N
Definite	41	25	34	146
Indefinite	55	28	18	120

The four tables show remarkably similar hierarchies comparing definite to indefinite relative clauses:

- for all four linguistic groups there is a higher rate of *wh*- forms in indefinite contexts than in definite ones.
- the zero variant occurs far more frequently in definite context than indefinite ones for all four groups.

Both of these results are in line with what has been found in earlier studies on native relative pronoun distribution (Van den Eynden, 1993, Tagliamonte, Smith and Lawrence, 2005) as they, too, had found that definite contexts favoured the use of the zero variant.

In terms of the factors of adjacency and definiteness, the overwhelming impression is again that the non-native groups are very similar to the native control group (and the native speakers considered in previous studies) and that they are following the variable rules of relative pronoun selection that had been uncovered in previous studies.

## 6.8 Multivariate analysis

Having examined the factors affecting the use of relative pronouns individually, this section will focus on the combined effect of the factors together in order to disentangle any interaction there might be and uncover which factors can be shown to play a significant role in the occurrence of the variants.

To establish what the variable rules of each linguistic group are, a separate multivariate analysis has to be conducted on each group before we can compare the four analyses together. This will allow us to examine whether they share hierarchies of constraint and ranges and ultimately to determine how similar their variable rules are.

I have decided to follow Guy and Bayley's method for the division of relative pronoun types for my own multivariate analysis, this involves categorizing relatives into one of three groups; *wh*-clauses, *that* and zero. By considering *who* and *which* together we will be able to examine the factor of animacy across all the variants. Like them and rather differently from later studies such as Tagliamonte, Smith and Lawrence (2005), I will be considering subject and object relative clauses together in the runs examining *wh*- and *that*. This will allow me to determine whether the non-native groups share the same constraints in terms of syntax as the native group for the variants which are used both with subject and object relative clauses. When we consider the factors affecting the use of zero we will only take object relative clauses into account as zero relative pronouns do not occur in subject relatives in Standard English.

The constraints governing the selection of *wh*- forms, *that* and zero forms are considered in the three tables below, each table presenting four multivariate analyses, one for each linguistic group. Each variant will be discussed on its own and section 6.9 will focus on an overall discussion into how native variability compares to what was found for the non-native speakers.

Table 6.22 below presents four independent multivariate analyses of factors contributing to *wh*- forms. Syntax of relative clause, animacy, adjacency and definiteness were considered in all four analyses. As noted earlier, only two factors were run for adjacency; adjacent and non-adjacent. In a multivariate analysis, the distribution of factor groups has to be 'orthogonal' (Guy, 1988:126-127), 'this means that the factor groups should be independent of each other' (Tagliamonte, 2006:181).<sup>99</sup> Although some degree of overlap

---

<sup>99</sup> An example given by Tagliamonte (2006:182) of non-orthogonality is a case where none of the oldest group of speakers in her sample were highly educated. Education was not considered in her analysis.

(i.e. interaction between groups) is acceptable, ‘some distortion of the results is probably occurring’ if this overlap is above 90% (Guy, 1988:131). Earlier it was noted that there were very few human object tokens in the data and cross-tabulations of the two factors revealed that the overlap between them was above 90% in all four linguistic groups. Consequently, the analyses below will combine the factors of syntax and animacy, giving us a single factor group that considers human subject, non-human subject and object relatives separately.<sup>100</sup>

Table 6.22: Four independent multivariate analyses into the probability of *wh*-

	English			French			German			Italian		
Input	.41			.50			.59			.52		
	FW	%	N	FW	%	N	FW	%	N	FW	%	N
		47	266		49	149		58	109		50	289
<b>Syntax and animacy</b>												
Subj/human	.90	87	126	.94	94	63	.84	88	51	.92	92	129
Subj/non-human	.37	28	46	.35	36	28	.58	67	21	.35	35	49
Object	.06	4	94	.07	7	58	.08	11	37	.07	8	110
Range	84			87			76			85		
<b>Adjacency</b>												
Adjacent	[.50]	46	239	[.46]	46	128	[.44]	55	92	.55	51	237
Non-adjacent	[.51]	56	27	[.74]	67	21	[.78]	71	17	.27	44	52
Range										28		
<b>Definiteness</b>												
Definite article	[.47]	41	146	[.42]	37	71	[.45]	49	65	.41	41	193
Indefinite article	[.54]	55	120	[.57]	60	78	[.57]	71	44	.67	68	96
Range										26		

Before proceeding, I will reiterate how the results are going to be interpreted. Examining the results factor by factor will allow us to determine which factors favour the use of *wh*- (those with weights over .50) and which disfavour *wh*- (those with weights under .50). The input represents the ‘overall tendency of rule application’ (Tagliamonte, 2006:156).

<sup>100</sup> The log-likelihoods of the runs with the combined factor groups were found to provide a better fit to the data.

In the table above, the inputs show that all the groups but the English speakers favour *wh*-forms as they are above .5. Having examined the factors individually, we will then use the ‘three lines of evidence (statistical significance, constraint ranking and relative strength of factors)’ (Tagliamonte, 2006:245) to establish whether the four groups share the same patterns. The factor groups which were found not to be statistically significant are given in square brackets.

**Syntax and animacy** is the only significant factor for three of the groups and the most significant factor for the Italian speakers. All four groups show very high ranges, which underlines how important this factor is in the conditioning of *wh*-. The constraint ranking is identical in the four groups, with human subject tokens very strongly favouring *wh*-forms, followed by non-human subject tokens and finally object tokens which very strongly disfavour *wh*- forms. Non-human subject tokens favour *wh*- only in the German multivariate analysis; this is likely due to the higher overall rate of *wh*- in their data. As the two factors were combined, our results are not completely comparable to previous studies but to a great extent they follow the expected constraint ranking; subjects favour *wh*- more than objects and animates favour it more than inanimates (Guy and Bayley, 1995). The non-native groups are very similar to the native group in terms of this factor.

The other factors considered in this analysis are not as important, but nevertheless show further similarities between the native and non-native groups. These factors were not found to be significant for the English, French and German groups, while they were for the Italians; this is likely due to the fact that the Italian speakers had a higher number of overall tokens. It is by examining the constraint ranking, alongside the aspects which we have just considered, that we will be able to establish if the non-natives are similar to the native speakers.

In terms of **adjacency**, previous studies had found that non-adjacent tokens were more likely to be used with *wh*- forms than adjacent tokens (Guy and Bayley, 1999:154). This is borne out by the English, French and German groups, where non-adjacent tokens

favour *wh*-, while adjacent tokens do not.<sup>101</sup> This factor was not found to be statistically significant for any of them however. The run considering the Italian speakers, on the other hand, found this factor to be significant and that the distribution was reversed; adjacent favours *wh*-, while non-adjacent disfavours it.

Although **definiteness** was only found to be significant for the Italian speakers for the selection of *wh*-, the hierarchies are identical in all four groups; as had been predicted by Quirk et al. (1985), for example, indefinite articles are more likely to be found with *wh*-relative clauses than definite articles.

How do the groups compare overall in terms of *wh*-? Although two factor groups are found to be significant only for the Italian speakers, the factor group combining syntax and animacy was significant in all groups. Moreover, it shows very high ranges in all four groups and an identical constraint ranking (subj/human>subject/non-human>object). The hierarchy is also the same in terms of definiteness; *wh*- is favoured more with indefinite than with definite. The factor of adjacency shows differences between the Italian speakers and the other three groups; unlike the other groups, Italian speakers favour *wh*- in adjacent contexts. This is the main difference between the groups. The overall impression is then one of similarity; the non-native groups share the patterns of the native groups.

Turning now to *that*, table 6.23 considers the factors found to be significant in its selection.

Table 6.23: Four independent multivariate analyses into the probability of *that*

	English			French			German			Italian		
Input	.24			.19			.11			.20		
	FW	%	N	FW	%	N	FW	%	N	FW	%	N
		26	266		24	149		14	109		26	289
<b>Syntax and animacy</b>												
Subj/human	<b>.34</b>	14	126	<b>.23</b>	6	63	<b>.51</b>	12	51	<b>.23</b>	9	129
Subj/non-human	<b>.89</b>	72	46	<b>.89</b>	64	28	<b>.80</b>	33	21	<b>.87</b>	65	49
object	<b>.47</b>	21	94	<b>.58</b>	24	58	<b>.31</b>	5	37	<b>.65</b>	29	110

<sup>101</sup> The range between the two factors is very low for the English speakers.



Range 58				66			49			64		
<b>Adjacency</b>												
Adjacent	[.49]	26	239	[.50]	23	128	[.51]	14	92	<b>.43</b>	21	237
Non-adjacent	[.55]	30	27	[.53]	33	21	[.46]	12	17	<b>.80</b>	52	52
Range										37		
<b>Definiteness</b>												
Definite	[.51]	25	146	[.44]	21	71	[.44]	9	65	[.50]	27	193
Indefinite	[.49]	28	120	[.56]	27	78	[.59]	21	44	[.51]	25	96
Range												

The inputs of the four groups show that *that* is quite strongly disfavoured; this is to be expected given the high rates of *who* found in all the speaker groups. As for *wh-*, the combined factor of syntax and animacy is the most important one for all four groups of speakers.

Again, only one factor group is statistically significant in all four linguistic groups; the one combining syntax and animacy. The German groups have a different constraint ranking than the three other groups, however. The native speakers and the Italian and French groups reveal that human subject contexts are those most disfavoured, while non-human subject are most favoured with *that* (subj/non-human > object > subj/human). The German speakers favour non-human subject most strongly but this factor is then followed by human subject and not object contexts (subj/non-human > subj/human > object).

**Adjacency** is statistically significant only for the Italian speakers. The expected hierarchy, whereby non-adjacent contexts are favoured with *that*, is found for the English, French and Italian groups. The German speakers have the opposite hierarchy.

Although **definiteness** was not found to be significant in any of the four languages and the factor weights in the English and Italian groups are very close to each other (slightly above or below .50). The French and the German groups favour *that* in indefinite contexts.

As with *wh-*, the patterns of the native English group are closely matched by the non-native groups in terms of *that*, demonstrating again that the non-native speakers mirror



In object relative contexts, the zero form is selected most frequently; it is highly favoured in the input of all four groups and represents between 63% and 84% of the object relative tokens.

Animacy was not found to be significant in the two groups where it was run and the constraint ranking is not shared. The English speakers favour animates more strongly than inanimates, while this hierarchy is reversed in the Italian group. Because of very low tokens for human objects, we cannot establish what the pattern of the French and German speakers might have been.

**Adjacency** was found to be statistically significant for the Italian speakers but not for the English speakers. All four groups seem to share the same constraint ranking, however, as adjacent tokens favour zero and non-adjacent tokens disfavour it. This follows the results of Guy and Bayley (1999:154).

In terms of **definiteness**, the findings match what had been found by Van den Eynden (1993:113); in all groups, the zero form is favoured in contexts where there is a definite article and disfavoured when an indefinite article is present. This factor is only statistically significant for the French and German speakers however.

Overall, the groups are not as similar in terms of zero relative pronouns as for *wh*- and *that*. This is due primarily to the fact that in considering only object relatives the runs for zero had a considerably lower number of tokens and the results are less conclusive. There are a number of similarities between the groups, however.

All four groups have very high percentages of zero in object relatives and share the same hierarchy for definiteness and for adjacency, and although the hierarchies differ in terms of animacy, it is most likely due to a very low number of human object relative tokens. Although we cannot establish with absolute certainty that the non-native groups share the patterns of the native speakers, for the most part the patterns have been acquired.

## 6.9 Discussion

The multivariate analyses for the three relative pronoun variants show a number of similarities and differences between the groups, both in terms of native/non-native but also across the non-native groups. The overwhelming consensus, however, is that the non-native groups are remarkably similar to the native speakers; although there are some differences in terms of constraint ranking and in terms of magnitude of effect, overall the non-native speakers appear to be following the strategies of the native e-mailers or those found in other native speakers in other studies. The differences can generally be explained in part to uneven data distribution so may not show diverging patterns at all. The multivariate analyses considering *wh*- and *that* show far more similarity between the groups than the ones for zero; they also have a much higher overall number of tokens.

It is not only in terms of the factors which highly restrict relative pronoun selection that the non-natives match the natives, but also for the factors that are largely unconscious, such as adjacency and definiteness. By and large, the non-native speakers have the same constraint ranking for these as native speakers.

This feature is a case where the non-native speakers have acquired the variability of native speakers, so case A of the categories discussed in chapter 4. The non-native speakers' use of relative pronouns is for the most part nearly identical to native speakers'. This similarity between the native and non-native speakers groups is not only on the surface, but also in terms of the underlying variability as well. Not only have the prescriptive rules affecting the distribution been acquired, but the variable constraints have as well.

Case A	French, German and Italian present similar variation patterns to native speakers	variation acquired
--------	--	--------------------

### **6.10 Conclusion**

As well as showing that the native English constraints governing relative pronoun selection have been acquired by the three non-native speaker groups, this chapter has also been able to demonstrate that differences in terms of syntactic structures of the source and target languages do not necessarily inhibit native-like acquisition of variable rules.

Furthermore, the fact that much of the variability was not taught to the non-native speakers was found to have no effect on their variable patterns as they match the native speakers. This is considerably different from what was found in terms of the future tenses which chapter 5 showed were not acquired in a native English-like manner by the non-native speakers. Here the non-native speakers were able to deal with the variability present in native English. The prescriptive aspects of variation might have contributed to the native-like patterns, as it reduced some of the potential variability.

By focussing on the target-based variability found in relative clauses rather than looking solely at the interference-based variability, this chapter has been able to uncover the extent to which unconscious variable rules may be acquired by non-native speakers.

## 7. Complementizers

### 7.1 Introduction

The fourth linguistic variable I will consider in this thesis is the use of *that* or zero complementizers as in (1) and (2):

- (1) I hope Ø you enjoyed the day and liked the city and the bears (b, Italian, e-mail)
- (2) I hope **that** in the future in all Switzerland we'll have some common projects at national level (f, French, e-mail)

Complementizer variation has been the subject of both synchronic (Tagliamonte and Smith, 2005, Thomson and Mulac, 1991, Torres-Cacoullous and Walker, 2003, Elness, 1984) and diachronic (Warner, 1982, Rissanen, 1991) research on native speaker English.

These studies provide an excellent backdrop from which to gauge the use of this variable in non-native speech, specifically with regard to frequencies and constraints on the use of the zero form.

There are a number of reasons why this feature will be useful to gain an understanding of the non-native acquisition of variation; first of all, unlike some of the other features considered in this thesis, such as relative pronouns where the variants were constrained by animacy, both forms are fully acceptable to native speakers in practically all circumstances. Hence, this variable is not given any focus in teaching, so if the non-native speakers match the native speaker patterns, then it is purely due to their ability to acquire native patterns unconsciously. Secondly, as will be discussed more fully in section 7.3, two of the source languages do not have a zero complementizer form similar to the English form, which means that any similarities between the native and non-native groups cannot be due to interference of the source languages.

This chapter will analyze the patterns of complementizer use found in the non-native speaker groups and attempt to determine whether the Swiss speakers have acquired the

variable rules of English speakers. The factors affecting complementizer use in native speakers will be presented in sections 7.4 and 7.5.

## 7.2 Complementizers

Syntactic structures, such as those introduced with verbs such as *think*, *say*, *mean*, can either have a *that* complementizer between the verb and the following clause or zero (as in examples 1 and 2 above).

Previous studies have reported two main findings. First of all, the use of the zero form has grown through the history of English. Although the zero form was rarely used in Old and Middle English, its use has increased over time, achieving a near categorical use in the twentieth century for some specific verbs (Tagliamonte and Smith, 2005:301, Thomson and Mulac, 1991:244). Secondly, previous studies have found that, in present day English style exerts a considerable effect on the selection of zero complementizer forms. Elsness (1984:521) found that in formal writing zero complementizers are used far less than in informal writing.

Looking at informal oral data, Thomson and Mulac (1991:242) found that overall the zero complementizer was used at a rate of 86%, while Tagliamonte and Smith (2005:300) found a rate of 84%.<sup>102</sup> Elsness (1984:521), looking at a subsection of written data from the Brown corpus, found rates of rates of 52% and 58% respectively in his two informal text categories (Press Releases and Fiction: Adventure & Western) and far lower rates of 15% and 1.3% in his two formal text categories (Belles Lettres & Biography and Learned & Scientific Writing) (see Kucera and Francis, 1967 and Ellegard, 1978 for further information about the Brown corpus and the Syntax Data Corpus which is a subset of it).

The non-native speakers of all three linguistic backgrounds use both variants, as can be seen from examples 3-8 below.

---

<sup>102</sup> Tokens such as *I think*, *you know* and *I mean* which categorically selected the zero complementizer were excluded in Tagliamonte and Smith's results.

### French speakers

- (3) I think **that** there's a virus in the past document I send to you. (f, French, e-mail)
- (4) I will take some copies [...], because I think **Ø** there is a virus (f, French, e-mail)

### Italian speakers

- (5) I guess **that** there will not be very many new people (b, Italian, e-mail)
- (6) I guess **Ø** you're back in Switzerland! (b, Italian, e-mail)

### German speakers

- (7) So I also think **that** we need to think carefully about the division of expenses. (h, German, e-mail).
- (8) I think **Ø** it is very important to have such a useful booklet (h, German, e-mail)

In terms of the teaching of complementizer forms to Swiss speakers, a survey of the grammar books reveals that the fact that there are two variants is never explicitly made clear. The two forms are used, however, both in the grammar books (Soars and Soars, 1987, Spencer, 1999) and other teaching materials<sup>103</sup> the Swiss students use and in the speech of their teachers. For example, a quick examination of the Swiss teacher data collected by Lukas Rosenberger as part of the Fonds National Project revealed that this was indeed the case, as there are cases of non-native Swiss English teachers using both variants. Somewhat similarly to the relative pronouns, if Swiss speakers have the same patterns as native speakers, it might be hypothesized that they acquired them subconsciously and not through overt and conscious teaching.

## 7.3 Complementizers in the non-native languages

As in the case of relative clauses, French and Italian do not have a zero complementizer variant. Complementizer forms in these two languages are somewhat similar to relative pronouns, in that the complementizer particle is also the more frequent of the possible relative pronouns (that is to say *que* for French, and *che* for Italian) (examples 9-10).

<sup>103</sup> Much of the literature read by students of English would have contained instances of both *that* and zero complementizer.



German and Swiss German, however, have both overt and zero complementizer forms (examples 11-12). Moreover, similarly to English, the *dass* form is seen to be more formal than the zero form. The variable patterns of complementizers have not been studied in German, so we cannot know if some of the other factors found to be significant in English complementizer use operate in German as well.

- (9) Je pense **que** tu as presque fini.
- (10) Penso **che** hai quasi finito.
- (11) a. Ich glaube, **dass** du schon fertig bist.  
b. Ich glaube **Ø** du bist schon fertig
- (12) d Ruth glaubt, **dass/ Ø** d Susann het s gmacht (Penner and Bader, 1995:103)

## 7.4 Previous studies

Studies which have focused on present day varieties of English (Elsness, 1984, Tagliamonte and Smith, 2005, Thompson and Mulac, 1991) can help establish how close the non-native speakers are to native norms. Although studies which focus on the diachronic side of complementizer variation (Warner, 1982, Rissanen, 1991) are useful in terms of establishing the factors that might influence variability, they cannot provide a suitable comparison to the non-native speakers.

Tagliamonte and Smith (2005) provide an in-depth presentation and summary of earlier studies and attempt to establish which of the factors mentioned in previous studies are relevant to their own data. Their analysis focused on the patterns of zero complementizer use of the oldest generation of speakers in several relatively isolated northern British communities (in Cumbria, Lowland Scotland and Northern Ireland). They found very high percentages of the zero complementizer (around 90%) and established that the zero form had become nearly categorical in some contexts (such as *I think, I mean*) (Tagliamonte and Smith, 2005:299). In the variable contexts of use, they also uncovered a number of internal factors which conditioned the use of the variants; 1<sup>st</sup> and 2<sup>nd</sup> person

subjects vs. other subjects (examples 13-14), present tense vs. past tense (examples 15-16), additional elements in the verb phrase (example 17) and finally adverbials between the verb phrase and complementizer (example 18) (Tagliamonte and Smith, 2005:301).

- (13) I guess  $\emptyset$  I will be the only one from Bern who will join the meeting. (m, German, e-mail)
- (14) he explain **that** there will be a meeting in Geneva (f, French, e-mail)
- (15) I think  $\emptyset$  it would be a great opportunity also to show what is going on in the other cities (b, Italian, e-mail)
- (16) But he told us, **that** he can speak German! (r, German, e-mail)
- (17) I don't think  $\emptyset$  I'm talking about that girl from Catalan country (h, German, e-mail)
- (18) I really hope **that** everybody is feeling fine. (a, Italian, e-mail)

Formality also exerts a considerable effect on the use of the complementizer variants. Tagliamonte and Smith (2005:300) and Thomson and Mulac (1991:242) found very high rates of zero complementizer in their analyses which considered varieties of informal spoken English. On the other hand, Elsness (1984:521), who examined written data, found low rates of the zero form in the more formal texts. He found higher rates of the zero form in the less formal texts but these rates were still rather lower than the oral data in Tagliamonte and Smith (2005) and Thomson and Mulac (1991).

These effects of formality and of register will need to be taken into account when considering the e-mail data in this analysis. The place of e-mail on the continuum between oral and written data is debatable (see Herring, 2001) so we cannot know *a priori* if the rates for the zero complementizer in our e-mail data will be high, as in spoken English, or low, as in formal written English. The native e-mail control data will play a crucial role here as it is directly comparable to the non-native data. If the native data shows high levels of zero complementizer then we would expect high levels from the non-natives as well, if we are to prove that the non-native speakers pattern like native

speakers. On the other hand, if the native data shows low levels, then we would expect the same from the non-natives.

## 7.5 Extraction and Coding

The present analysis will focus on comparing the two e-mail corpora and not consider the oral non-native data because, as shown above, there are considerable differences between oral and written uses of complementizers. It will thus be more revealing to establish whether there are any comparable patterns between the native and non-native speaker groups. Because only the e-mail data is considered, my results will not be strictly comparable to Tagliamonte and Smith (2005) and Thomson and Mulac (2001) as they dealt with speech. Every instance where either *that* or *zero* could have been used were extracted from the data.<sup>104</sup>

The factors which were coded in this analysis and which will be studied in detail are very similar to those examined by Tagliamonte and Smith in their work. As well as **speaker** and **speaker's native language**, which are the external factors in this analysis, I will focus on **subject of the matrix clause**, the **tense of the verb**, whether there are any **additional elements to the Verb Phrase**, whether the Verb phrase and the complementizer are **separated by adverbs or adverbials** and finally the **lexical verb** which the complementizer follows.

The **subject of the matrix clause** was found to be important in the research both of Thomson and Mulac (1991:242) and Tagliamonte and Smith (2005:294), where it was found that first and second person subjects have a higher proportion of zero complementizers than third person subjects. The tokens have been coded for whether they are first person singular (19), first person plural (20), first person singular but with the pronoun omitted (21), second person (singular or plural) (22), third person singular pronoun, *he* or *she* (23), third person singular pronoun *it* (24), third person plural pronoun (25), third person singular Noun Phrase (26), third person plural Noun Phrase (27), no

---

<sup>104</sup> Rather differently from Tagliamonte and Smith (2005), there were no cases of parentheticals which were in a position where the *that* form was not a possible option (i.e. – ‘She’s very nice, I think’).

overt subject (28). Because of a low number of tokens for some of these factors, all the first person subjects are combined into a single group and the same was done with the various third person pronouns. In the multivariate analysis, all non first or second person subjects are combined regardless of whether they are pronouns or noun phrases. This was possible because these subjects had been shown to pattern in a similar manner.

- (19) I guess **that** I still the NORE<sup>105</sup> for this year (l, French, e-mail)
- (20) We knew **that** X didn't speak German (o, German, e-mail)
- (21) Ø hope Ø you're all doing fine (h, German, e-mail)
- (22) Do you also think **that** we should buy a Firewall program? (c, Italian, e-mail)
- (23) But he told us, **that** he can speak German! (r, German, e-mail)
- (24) it's about time **that** things get clear in this meeting story (&, French, e-mail)
- (25) they told me **that** they can give us some sample materials (b, Italian, e-mail)
- (26) Young teenagers (girls), think Ø their physical appearance is very important (f, French, e-mail)
- (27) If everybody shows me **that** we're all a big team (c, Italian, e-mail)
- (28) P.S.: I'm sorry, today we had so much sun, **that** my last two neurons got xx (c, Italian, e-mail)

The **tense of the matrix verb** was coded with a three way distinction, verbs in the present tense (29), verbs in the past tense (30), and sentences with no verb (31). Tagliamonte and Smith (2005:304) had found that verbs in the present tense favoured the zero complementizer more than past tense verbs. There are relatively few tokens of sentences with no verb, so these will be excluded in the multivariate analysis.

- (29) I think also **that** X is in Geneva, isn't it? (w, French, e-mail)
- (30) In the meantime I found **that** he wrote his notes for fundraising on the IFMSA web page (c, Italian, e-mail)

---

<sup>105</sup> NORE = National Officer for Research Exchanges

- (31)        The fact **that** the computer is not always on minimizes greatly the possibilities for anyone to access it (j, French, e-mail)

In terms of **additional elements in the verb phrase**, the tokens were coded for whether there were no additional elements (32), whether the additional element was a modal (33) or whether the additional element was a negation form (34). The tokens which had both a modal and a negation form were given a separate code (35). Tagliamonte and Smith (2005:304) found that ‘simpler constructions’ (i.e. those without additional elements in the verb phrase) favoured the zero complementizer.

- (32)        I think Ø you'll understand why (c, Italian, e-mail)
- (33)        To sum it up it can be said **that** SCOME-CH has to build a solid structure for concrete projects (f, French, e-mail)
- (34)        I don't think Ø it would be necessary to buy a multi-user license (j, French, e-mail)
- (35)        I cannot promise **that** I can attend (\*, French, e-mail)

Somewhat similarly to the previous factor, the tokens in the factor considering **other additional elements in the matrix clause** were coded for absence (36) or presence (37) of additional elements. Again, Tagliamonte and Smith (2005:304) found that tokens without additional elements favoured the zero complementizer.

- (36)        if you Ø thought **that** you had already won the portwine bottle (p, German, e-mail)
- (37)        I really hope **that** everyone arrived in Kopaonik as per travel-time-table! (a, Italian, e-mail)

In terms of **lexical verbs**, a number of specific verbs were analyzed to determine how they affected the variability; any verb occurring frequently enough to allow it to be analyzed on its own was considered (*think, hope, tell, say and know*). The cut-off point for this was at least 24 tokens in the whole of the IFMSA data set.

Except for *hope*, which was not examined in isolation by Tagliamonte and Smith (2005), the other specific lexical verbs are the same ones considered in Tagliamonte and Smith's 2005 study.

## 7.6 Results

### Overall distribution

The non-native speaker e-mail corpus yielded 576 tokens, and the native corpus 328. The breakdown of the tokens, by native language and by *that* or zero complementizer is provided in Table 7.1 below.<sup>106</sup>

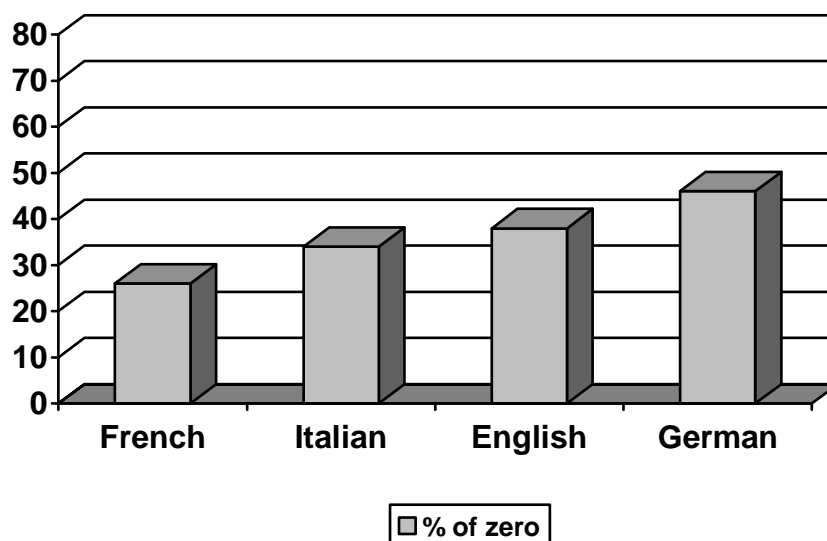
Table 7.1. Overall distribution of zero use in complementizers (e-mails).			
	% of zero	% of <i>that</i>	Total N
IFMSA	33	67	576
French	26	74	197
German	46	54	87
Italian	34	66	292
Natives	38	62	328

The occurrence of the zero complementizer form is far lower in all four groups than what was found by Tagliamonte and Smith (2005) and by Thomson and Mulac (1991), as their research had found near categorical rates of zero in certain contexts. This is not entirely surprising, however, as it had been noted in the research of both groups that speech is more likely to show the zero form than written data. The rates found in the present analysis are at a mid-point between the informal and formal texts that Elsness (1984) had considered. This further underlines how e-mails are a separate medium from both oral and written data and why it is crucial to use a native English control group of e-mailers to compare to the Swiss data. It might have been hypothesized that the non-native speakers would have much higher rates of *that* given their mother tongues do not have forms comparable to the English zero form and the fact that they are not explicitly taught but this is not the case at all. The native group is not substantially different from the three

<sup>106</sup> In this table and the other tables in this chapter, the native data comes solely from my own e-mail corpus.

non-native groups in terms of percentage of zero complementizer forms, being at a mid-point between the German speakers and the French and Italian speakers as can be seen from graph 7.1 below.

**7.1 Percentage of zero complementizer**



A chi square test considering all four linguistic groups reveals that the differences between them are significant,<sup>107</sup> so, at this stage, we cannot state that the four groups are identical in their complementizer use. The three non-native groups make use of both variants, however. Their usage is variable, something which was not the case for the future variants, for example. By examining the distribution of the variants across the various factor groups in more detail, we may be able to better establish how similar the non-native groups are to the native speakers.

Note that the German speakers use more of the zero complementizer than the native group; this may be due to interactions with the lexical verb. If the German speakers have a higher proportion of *think* than the other groups, for example, then this could explain the difference in zero complementizer use. We shall turn to lexical verb in a moment, but

<sup>107</sup> (df=3,  $\chi^2 = 12.1882144658936$ ,  $p < 0.01$ .)

if this factor cannot explain the differences found in the German speakers then we will need to come back to this higher rate of use.

In terms of **individual speaker** use, some of the IFMSA members provide a far higher number of tokens than others; what is crucial, however, is all the speakers are variable to some extent, so there is no risk that one member might be skewing the results of one of the language groups. In terms of speakers per individual linguistic group, there are 14 different French language speakers who have provided tokens of complementizers, 15 German language speakers and 6 Italian language speakers.

Because, as was the case for relative pronoun selection, there are a number of factors which highly constrain the variability, an in-depth examination of these factors will allow us to gain a greater perspective into any similarities or differences between the linguistic groups.

### **Lexical verb**

The first factor considered separately was the specific lexical verb which preceded the complementizer clause. Both Tagliamonte and Smith (2005) and Thompson and Mulac (1991) found that some verbs (namely *think*, *mean*, *know*) showed far higher rates of zero complementizer than other verbs. This difference between specific verbs was explained in two ways; first of all, some verbs, such as *think*, were more likely to have an epistemic meaning and more likely to use the zero form, and secondly the higher frequency verbs were also more likely to show high rates of zero complementizer.

The present analysis will consider *think*, *hope*, *tell*, *say* and *know*, as these were the verbs which occurred most frequently in the non-native data. The other verbs which were found to be more likely to use the zero complementizer did not occur frequently enough to warrant being considered individually.

### Think



The two e-mail corpora examined in this thesis provided around 200 tokens of *think*. This represents more than 20% of the overall tokens considered in this analysis, so *think* is one of the most frequently occurring verbs, as was the case for Tagliamonte and Smith (2005) and for Thomson and Mulac (1991).

The percentages of *think* with a zero complementizer are lower for all four groups than what was found by Tagliamonte and Smith (2005:301) (93% after the tokens of *I think* had been removed) and by Thomson and Mulac (1991) (Table 7.2). Nonetheless, the native speaker group is far closer to this near categorical average (with 90% zero) than the three non-native groups (with ranges between 50 and 70%). Note that *think* has far higher rates of zero complementizer than the overall distribution in all four groups. Although the non-native groups have lower rates than the natives, they share the direction of effect.

Table 7.2: Distribution of complementizer forms for *think*

	% of zero	% of <i>that</i>	Total Ns
IFMSA	<b>52</b>	<b>48</b>	161
French	<b>51</b>	<b>49</b>	51
German	<b>68</b>	<b>32</b>	22
Italian	<b>54</b>	<b>46</b>	80
Natives	<b>90</b>	<b>10</b>	39

Although the non-native speakers show variation in terms of this lexeme, it is at a far lower rate than found in native speakers. While it will only be through an analysis of the multivariate analysis of this feature that we will be able to establish for certain whether the non-native groups are truly following different patterns, these results nevertheless signal that there are some differences between the native and non-native groups.<sup>108</sup> The lower rates found in the non-native speakers might also show that their e-mails are generally more formal than natives; recall that for the relative pronouns, only the non-native speakers had tokens of the highly formal variant *whom*. I will return to this point in the discussion of the feature.

<sup>108</sup> A chi square calculation of the four groups confirms this is a significant result: chi-square = 18.0340485792825,  $p < 0.001$ .

Despite the difference with the native group, the three non-native groups show very similar rates; there is no significant difference between them.<sup>109</sup> This is an important point to consider as it may turn out that complementizer selection is a case which can help posit a case for a single variety of English used by all three linguistic groups.

### Hope

*Hope* occurred frequently enough in the e-mail data for it to be considered individually, as there were nearly 100 tokens for the non-native speakers and 25 in the native e-mails. This verb was not considered separately in other studies, so we do not know what the native oral rates for it are.

As in the case of *think*, *hope* demonstrates a high proportion of the zero complementizer in all four groups, with the German speakers being closest to the native percentages and the French speakers furthest away (Table 7.3).

Table 7.3: Distribution of complementizer forms for *hope*

	% of zero	% of <i>that</i>	Total N
IFMSA	<b>62</b>	<b>38</b>	95
French	<b>56</b>	<b>44</b>	27
German	<b>81</b>	<b>19</b>	16
Italian	<b>60</b>	<b>40</b>	52
Natives	<b>88</b>	<b>12</b>	25

Here again, although all the groups vary and show percentages of zero complementizer forms above fifty percent, there are differences between the groups. In this instance, the German speakers are very close to the native speakers while the other two groups delete the complementizer about 20% less. The low number of tokens for the German speakers makes it difficult to be sure this is significant. The differences between the four groups are significant but, as for *think*, when the only three non-native groups are considered, the difference is found not to be significant.<sup>110</sup>

<sup>109</sup> A chi square calculation of the three non-native groups is found to be not significant: chi-square = 1.9266286702428,  $p < 1$ .

<sup>110</sup> All four groups: chi square= 9.3976,  $p < 0.025$ . Three non-native groups: chi square = 3.1209,  $p < 1$ .

## Tell

The next high frequency lexeme is *tell*. Tagliamonte and Smith (2005:301) had found rates of 64% percent of the zero form with *tell* in their data, which was considerably lower than the other verbs they examined and lower than the overall distribution (which was 80%). Directly reported speech has to be separated from indirectly reported speech with *tell* because in direct speech there is no complementizer (examples 38-40).

- (38) He told me 'I'm happy' – direct speech
- (39) He told me **that/ Ø** I was happy - indirect speech
- (40) He told me **that/ Ø** he was happy. – indirect speech

While the number of tokens for *tell* is rather lower than for *think* and *hope*, the distributions of the various linguistic groups can still be analyzed, and once again there are considerable differences between the natives and the German speakers on one hand and the French and Italian speakers on the other (Table 7.4).<sup>111</sup>

Table 7.4 Distribution of complementizer forms for *tell*

	% of zero	% of <i>that</i>	Total N
IFMSA	<b>23</b>	<b>77</b>	31
French	<b>11</b>	<b>89</b>	9
German	<b>44</b>	<b>56</b>	9
Italian	<b>15</b>	<b>85</b>	13
Natives	<b>45</b>	<b>55</b>	11

## Say

Tagliamonte and Smith (2005:301) found that the zero complementizer occurred at a rate of 85% with *say* in their data. There are slightly over 40 tokens of it in the two e-mail corpora. The distribution of the tokens is not ideal as the French and native English speakers provide the majority of the tokens, so the results of the German and Italian speakers (with totals of 3 and 4 tokens respectively) cannot be considered to be truly indicative of the situation (table 7.5).

<sup>111</sup> The low number of tokens per cell means that it is not possible to establish whether these figures are significant.

Table 7.5 Distribution of complementizer forms for *say*

	% of zero	% of <i>that</i>	Total N
IFMSA	21	79	24
French	18	82	17
German	33	67	3
Italian	25	75	4
Natives	37	63	19

As was the case for *hope* and *tell*, *say* occurs at a higher rate with the zero complementizer for the German and native English speakers. The French rate (18%) is considerably lower than the native group (37%). The native e-mail rate (37%) is much lower than what was found in speech.

### Know

Similarly to *say*, Tagliamonte and Smith (2005:301) found that *know* occurred with the zero complementizer at a rate of 85%. There are around fifty tokens of *know* in the e-mail corpora, and in this case, it is the French and the German speakers that have a far lower number of tokens than the other two groups (Table 7.6).

Table 7.6 Distribution of complementizer forms for *know*

	% of zero	% of <i>that</i>	Total N
IFMSA	43	57	28
French	25	75	4
German	40	60	5
Italian	47	53	19
Natives	42	58	24

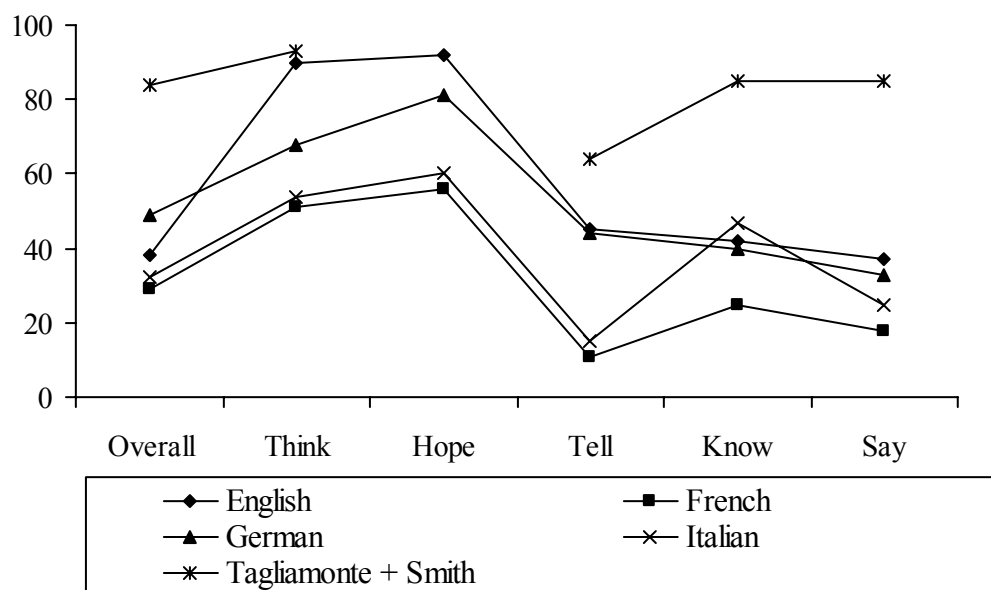
Table 7.7 presents a summary view of the percentages of zero complementizer for all the verbs studied individually, while figure 7.2 presents these results in graph form. The rates of the different verbs alongside each other must be considered as this will allow us to establish whether the groups share the same patterns despite having different overall distributions. Figure 7.2 also plots the results from Tagliamonte and Smith (2005:301) to allow us to see how the oral data compares to the e-mail data.

Table 7.7 Percentage of zero complementizer by verb

	Overall	Think	Hope	Tell	Know	Say
English	38	90	92	45	42	37
French	29	51	56	11	25	18
German	49	68	81	44	40	33
Italian	32	54	60	15	47	25

What is most striking when examining figure 7.2 is that the four e-mail groups show remarkably similar patterns; despite differences in percentages, the four e-mail groups have the highest rates for the same verbs. The hierarchy they all show is *hope* > *think* as the verbs with the highest rates of zero complementizer with *tell*, *know* and *say* showing lower rates. The French and Italian groups have patterns that are marginally more similar to each other than to the other two groups and the same holds for the English and German groups; nevertheless the overall picture is that the three non-native groups have similar patterns to the native control group.

Figure 7.2: Percentage of zero by verb



Because Tagliamonte and Smith (2005) did not provide the rate of zero complementizer use with *hope* in their results, there is a gap on figure 7.2. Although the rates of zero

complementizer found by them are considerably higher than for any of the e-mail groups, there are several points of similarity. *Tell* is the lexical verb with the lowest rate of zero complementizer in all five groups and *think* has one of the highest rates. Similarly to the French and Italian groups, Tagliamonte and Smith's results show that *know* is used considerably more with the zero complementizer than with *tell*.

Overall, in terms of specific verbs, these results provide us with two major findings. First of all, the native e-mails show a distribution similar to the findings in previous studies in terms of specific verbs. The rates of zero complementizer are slightly lower than what was found in spoken data, but the near categoricity of the zero form with *think* is replicated here. The e-mail medium did not affect the variable rules underlying the distribution of the zero complementizer. Secondly, although the three non-native groups have lower percentages than the native speaker control group, figure 7.2 demonstrated that their patterning is very similar to the native speakers. Their percentages fall between those of the formal and informal text categories that Elness (1984:521) had examined, again underlying how difficult it can be to place e-mail in terms of written or oral registers.

### **Subject of matrix clause**

Previous studies found that the subject of the matrix clause influenced complementizer choice in that first and second person subjects were more likely to occur with a zero form than third person forms. The data in the present study will be considered in terms of a four-way distinction with singular and plural subjects considered together; first person subjects, second person subjects, third person pronoun subjects and third person NPs. This follows the findings of previous studies.

The results of previous studies are matched by the non-native e-mailers and the native e-mailers as is demonstrated in Table 7.8 below. First person subjects have the highest rates

of the zero complementizer in all four groups and, except for the German speakers; these are followed by second person subjects.<sup>112</sup>

Table 7.8: Percentage of zero complementizer by subject of the matrix clause

	French		German		Italian		English	
	% of zero	Total N	% of zero	Total N	% of zero	Total N	% of zero	Total N
First person	<b>38</b>	119	<b>60</b>	60	<b>44</b>	192	<b>52</b>	143
Second person	<b>29</b>	7	<b>22</b>	9	<b>18</b>	22	<b>36</b>	75
Third pers. Pronoun	<b>4</b>	49	<b>25</b>	8	<b>16</b>	44	<b>18</b>	56
Noun Phrase	<b>18</b>	22	<b>0</b>	10	<b>6</b>	34	<b>22</b>	54
Total	<b>26</b>	197	<b>46</b>	87	<b>34</b>	292	<b>38</b>	328

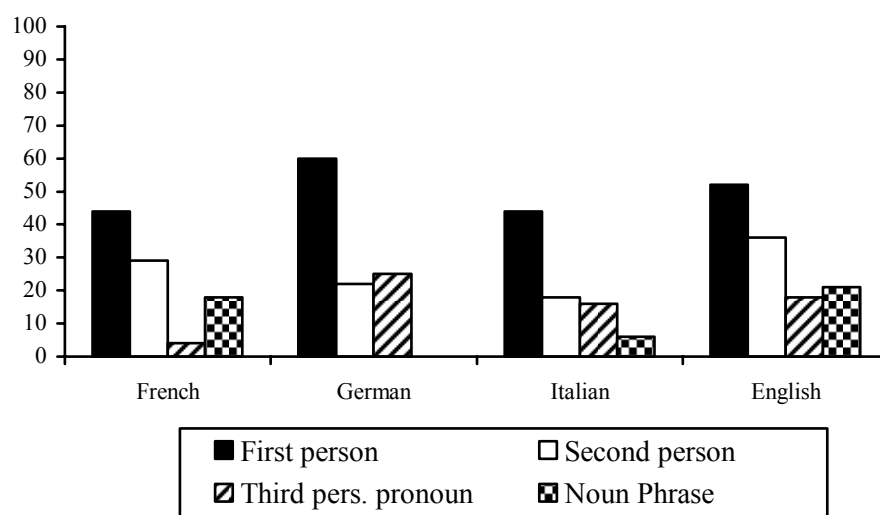
There are a number of differences in terms of the hierarchy in terms of the third person categories however; the English and the French speakers have higher percentages of zero with NPs than with third person pronouns while it is the opposite for the German and Italian speakers.

Figure 7.3 plots the percentages of zero by subject in a graph to make the distribution more apparent; despite some differences in terms of third person subjects, the four groups are quite similar.

The non-native groups share the hierarchies of the native speakers and are varying their use of the zero complementizer according to the subject of the matrix clause in a very similar way to the native e-mailers.

<sup>112</sup> Although Tagliamonte and Smith (2005) had found a considerable amount of interaction terms of first person subjects and *think*, this analysis did not uncover the same categorical distribution of *I think* with the zero form for any of the linguistic groups.

Figure 7.3: Percentage of zero by subject



## Tense

Previous studies (Tagliamonte and Smith, 2005, Thompson and Mulac, 1991) found that verbs in the present tense are more likely to be used with the zero complementizer form than past tense verbs. The results for the present study are shown in Table 7.9 and in figure 7.4 below.

Table 7.9: Percentage of zero complementizer by verb tense

	French		German		Italian		English	
	% of zero	Total N	% of zero	Total N	% of zero	Total N	% of zero	Total N
Present	30	162	49	65	36	262	38	269
Past	12	34	42	19	12	24	42	43
No verb	0	1	0	3	17	6	13	25



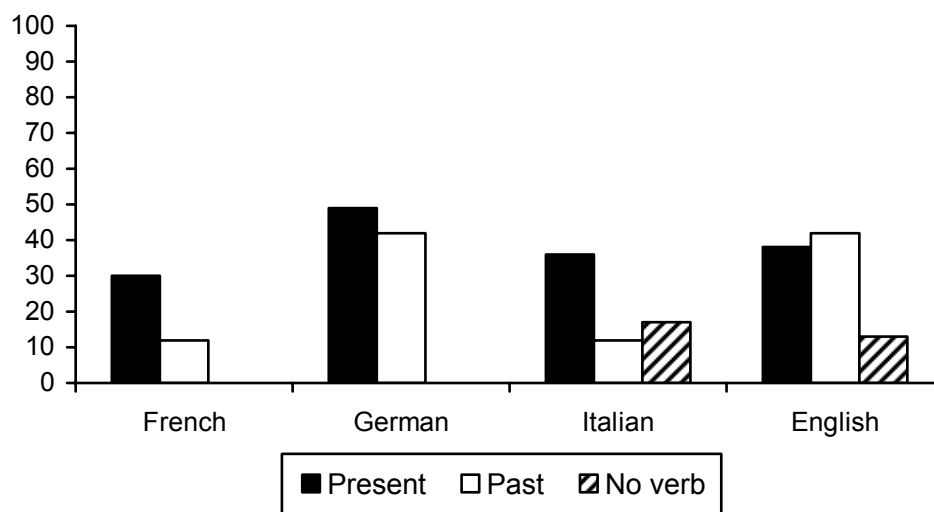
**Figure 7.4: Percentage of zero by verb tense**

Table 7.9 and figure 7.4 show that, unlike the factors of lexical verb and the subject of the matrix clause, the groups are very different in terms of verb tense. Rather unexpectedly the English group shows a higher proportion of zero deletion with past verb than with present ones, which is at odds with both the non-native e-mailers and with the findings for native speakers in previous studies. It may be because this analysis includes tokens of *I thought* while Tagliamonte and Smith's 2005 study did not. This discrepancy could also be linked to relatively low proportion of past tense forms opposed to the present tense tokens. The multivariate analysis should be able to determine if this is the case or whether there are other factors involved in this difference.

No conclusions about the tokens found in sentences without verbs can be drawn. Expect for the English natives who have 25 tokens with no verb, the number of tokens is far too low to enable us to establish whether this factor favours or disfavours the zero complementizers. As mentioned earlier, 'no verb' will not be considered in the factor of verb tense in the multivariate analysis.

### **Additional elements in Verb Phrase**

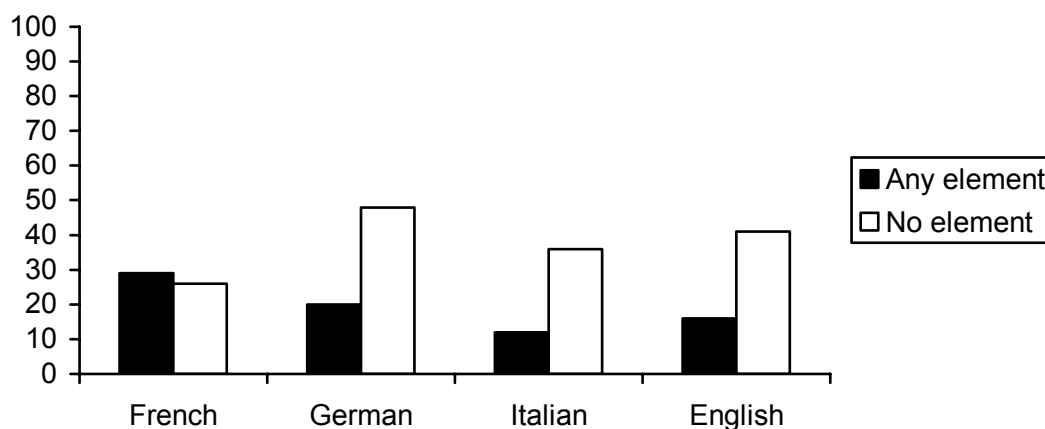
Previous studies found that the presence of modal and negation forms lowered the likelihood of the zero complementizer being used. The results presented in Table 7.10 and Figure 7.5 consider this in terms of the e-mail corpora.

Table 7.10 presents the percentages and figures of modal and negation forms separately while figure 7.5 collapses them together as there are too few tokens otherwise. This allows us to examine the rates of zero complementizer in terms of additional elements vs. no elements.

Table 7.10: Percentage of zero complementizer depending on elements in VP

	French		German		Italian		English	
	%	Total N	%	Total N	%	Total N	%	Total N
Negation	80	5	0	0	25	4	100	4
Modal	0	6	33	3	10	21	8	40
Modal and Negation	0	3	0	2	0	1	0	1
Any element	29	14	20	5	12	26	16	45
No Element	26	183	48	82	36	266	41	283

**Figure 7.5: Percentage of zero deletion by elements in VP**



Although the hierarchy is very similar for the German, Italian and English groups, in that they all, as predicted, have higher rates of zero complementizer in clauses without

additional elements, the overall low rate of tokens containing additional elements means that it is dangerous to attribute too much importance to these findings. The low number of tokens may also account for the divergent distribution of the French speakers.

### Additional elements in matrix clause

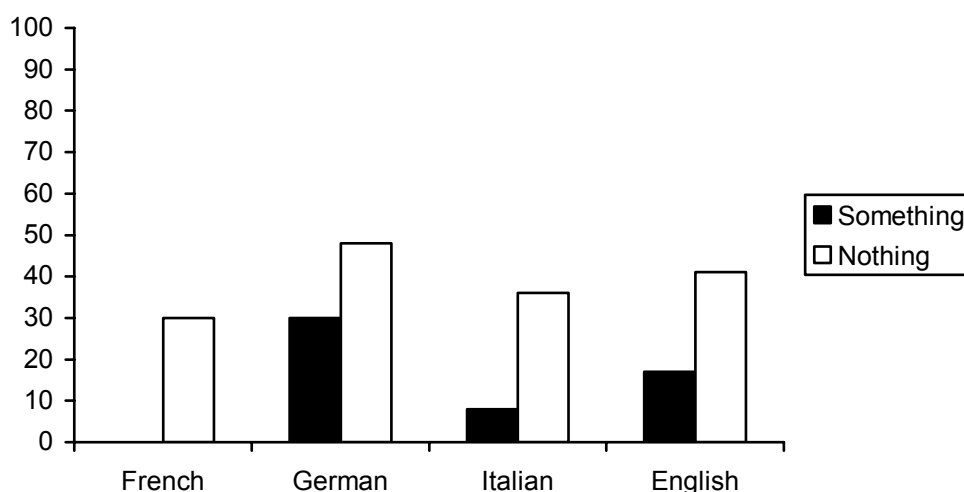
Tagliamonte and Smith's 2005 research found that additional elements in the matrix clause, such as adverbials, decreased the likelihood of a zero complementizer being used.

Table 7.11 and figure 7.6 below analyze the distribution in the e-mail data.

Table 7.11: Percentage of zero complementizer with additional elements in the matrix clause

	French		German		Italian		English	
	%	Total N	%	Total N	%	Total N	%	Total N
Something	0	23	30	10	8	24	17	48
Nothing	30	174	48	77	36	268	41	280

**Figure 7.6: Percentage of zero complementizer with additional elements in the matrix clause**



All four e-mail groups show the predicted distribution; additional elements in the matrix clause lower the use of the zero complementizer form.

Having examined the factors one by one, the next step is to consider the factors alongside one another in a multivariate analysis to determine with greater accuracy which factors

have a significant effect on the rates of the two complementizer forms. Before turning to the multivariate results, however, the main findings of each factor in terms of similarity between oral and written data (by comparing my results to previous studies) and between the native and non-native groups will briefly be summarized.

- **Overall rates:** the main differences in zero complementizer use are between the e-mail and the oral data. Whereas Tagliamonte and Smith found overall zero use to be around 84% in their data, the e-mail groups are all around 35%. The English and the German groups have higher rates of the zero form than the French and the Italian groups however. Although higher than the rather low rates (15% and 1%) Elsness (1984) had found in his two categories of formal texts, all four groups have lower overall rates than what he had found (52% and 58%) in two categories of informal texts (Elsness, 1984:521). In terms of complementizers, the register of e-mails is more formal than oral data and informal written data, but this affects native and non-native speakers alike.
- **Lexical verb:** Despite some differences in overall distribution of the zero form, the four e-mail groups have very similar patterns for the different verbs; *hope* and *think* are considerably higher than the other verbs. This matches the findings in Tagliamonte and Smith (2005).
- **Subject of matrix clause:** The four e-mail groups have the pattern found for oral data in Tagliamonte and Smith (2005); first person subjects have the highest rate of the zero variant followed by second person subjects then other subjects.
- **Tense:** This factor revealed some surprising findings; while the three non-native groups had higher rates of the zero form in present tense as was predicted by previous studies, the native e-mailers had the opposite.

- **Additional elements in Verb Phrase:** Here the order predicted was an increase of the zero complementizer in cases where there was no additional element in the VP; all but the French speakers showed the expected order.
- **Additional elements in Matrix clause:** As for the previous factor group, the absence of additional elements was found to favour the zero complementizer. This was found to be the case for all four e-mail groups.

Examining the whole range of factors influencing the use of complementizers, the overwhelming conclusion one can draw is that the non-native groups are very similar to the native e-mailers. The multivariate analysis will now be able to verify if it is indeed the case that the non-native groups have managed to acquire the native variable patterns of this feature.

Before turning to the multivariate analysis, the possibility that interference from the source languages affected the results needs to be considered. Recall that French and Italian did not have a form comparable to the English zero complementizer, while both High German and Swiss German did. Is this noticeable in the results? The German speakers have higher overall rates of zero complementizer than the two other non-native groups.

This may signal that the zero complementizer form in German helped this group acquire it in English. Before ascribing too much importance to this fact, however, remember that the German speakers had rates closer to native speakers than the other groups for the future variable as well. The higher rates found in the German group could be due to a higher level of English competency rather than structural similarities. Furthermore, more significant than mere percentages are the patterns which are matched not only by the German speakers, but by the Italian and French speakers as well. These patterns are those of English, not of German, so the German speakers would have had to acquire them, just as the French and Italian speakers did.

## 7.7 Multivariate analysis

Four separate multivariate analyses considering the probability of the zero complementizer being used were run; each analysis examined one of the linguistic groups in isolation. This was to allow us to establish what the variable rules and hierarchies were for each group and then compare them to one another. Table 7.12 below presents the results of the four analyses in a single table to allow for an easier comparison, however.

The five separate factor groups of lexical verb, subject of Matrix Clause, tense, additional elements in VP and additional elements in Matrix Clause were considered. Because of differences in terms of overall distribution in the different linguistic groups, some factors were excluded from one or more of the analyses:

- for **tense**, the tokens which been coded as having no tense were excluded in all the analyses as there were generally too few of them to allow for an accurate analysis.
- for **subject of the matrix clause**, all third person subjects, NP and pronoun, were merged together for the English and the Italian analyses, creating a three-way division. For the French and the German data, a two-way distinction was made considering first person subjects against all other subjects due to the low number of tokens of non-first person subjects.
- The factor of **additional elements in the matrix clause** was not run for the French group as all the tokens with additional elements had been found to have the *that* variant. Goldvarb cannot run a multivariate analysis on factor groups where one of variants is used 100%.
- The factor group examining **additional elements in the VP** was not considered for the German or the French speakers as there were too few tokens of complementizers with additional elements to make an analysis conclusive.
- for **lexical verb**, verbs with fewer than nine tokens were excluded from the individual analysis as their inclusion could skew the data. For the French speakers *know* was excluded, for the Italian speakers *say* was excluded and for the German speakers both *know* and *say* were excluded.

This will make the comparison across varieties marginally more complicated but we must remove any factor groups whose distribution is deemed to be linked solely to token numbers as this can obscure the actual patterns.

Table 7.12. Four independent multivariate analyses into the probability of zero complementizer

		English			French			German			Italian		
		%	Total N		%	Total N		%	Total N		%	Total N	
		38	328		26	197		46	87		34	292	
Input		.36			.20			.42			.27		
		FW	%	N	FW	%	N	FW	%	N	FW	%	N
Lexical Verb													
	Think	<b>.94</b>	90	39	<b>.81</b>	51	51	<b>.70</b>	68	22	<b>.75</b>	54	80
	Hope	<b>.90</b>	88	25	<b>.84</b>	56	27	<b>.81</b>	81	16	<b>.78</b>	60	52
	Know	<b>.58</b>	42	24	<b>n/a</b>	[25	4]	<b>n/a</b>	[40	5]	<b>.68</b>	47	19
	Say	<b>.55</b>	37	19	<b>.47</b>	18	17	<b>n/a</b>	[33	3]	<b>n/a</b>	[25	4]
	Tell	<b>.64</b>	45	11	<b>.34</b>	11	9	<b>.65</b>	44	9	<b>.31</b>	15	13
	Other	<b>.30</b>	21	210	<b>.23</b>	7	89	<b>.19</b>	16	32	<b>.22</b>	10	123
Range		64			61			62			56		
Subject of MC													
	1 <sup>st</sup>	[.52]	52	143	[.55]	37	118	<b>.65</b>	60	60	[.53]	44	192
	2 <sup>nd</sup>	[.60]	36	75							[.29]	18	22
	Other	[.41]	20	110	[.42]	10	79	<b>.20</b>	15	27	[.49]	12	78
Range								45					
Additional Elements in VP													
	Some	<b>.27</b>	16	45	<b>n/a</b>	29	14	<b>n/a</b>	20	5	[.34]	12	26
	None	<b>.54</b>	41	283	<b>n/a</b>	26	183	<b>n/a</b>	48	82	[.52]	36	266
Range		27											
Additional Elements in MC													
	Some	<b>.26</b>	17	48	<b>n/a</b>	0	23	<b>.17</b>	30	10	<b>.18</b>	8	24
	None	<b>.55</b>	41	280	<b>n/a</b>	30	174	<b>.55</b>	48	77	<b>.53</b>	36	268
Range		29						38			36		
Tense													
	Present	[.50]	38	269	[.51]	30	162	[.49]	49	65	[.52]	36	262
	Past	[.42]	42	43	[.47]	12	34	[.53]	42	19	[.27]	12	24
Range													

In terms of **input**, all four groups disfavour the zero complementizer; this is rather different from what had been found in Tagliamonte and Smith (2005). This is most likely due to the fact that the data in the present analysis comes from e-mails rather than speech.

**Lexical verb** was found to be significant for in all four variable rule analyses; moreover all four linguistic groups show similarly high ranges for this factor. In terms of the hierarchy of this factor, there are a few differences but the overall impression is that the groups pattern in a very similar fashion. *Think* and *hope* are the two most favoured forms overall and the category of ‘other verbs’ is the most disfavoured factor in all four groups. There are some differences in terms of the hierarchy of the other verbs studied, especially given that some verbs were excluded from some analyses, but their weights generally are between *think/hope* and ‘other’.

Table 7.13 below presents the differences in hierarchy between the various linguistic groups in terms of lexical verb.<sup>113</sup> The verbs given in parentheses are those which were not included in the multivariate analysis, their position in the table is related to their percentage of zero complementizer.

Table 7.13: Hierarchy of lexical verbs for zero complementizers

English	<b>Think</b>	> <b>hope</b>	> <b>tell</b>	> <b>know</b>	> <b>say</b>	> other
French	<b>Hope</b>	> <b>think</b>	(> know)	> say	> tell	> other
German	<b>Hope</b>	> <b>think</b>	> <b>tell</b>	(> know )	(> say)	> other
Italian	<b>Hope</b>	> <b>think</b>	> <b>know</b>	(> say)	> tell	> other

The **subject of the matrix clause** was found to be significant only for the German speakers. However, as predicted by Tagliamonte and Smith, first person subjects are favoured in all four groups. First person has the highest factor weight in the three non-native groups, but second person subjects have a higher rate for the native speakers. This is somewhat unexpected as it does not match Tagliamonte and Smith’s results.

<sup>113</sup> Verbs in bold are those which the multivariate analysis found to favour the use of the zero complementizer.



**Additional elements in the VP** was found to be significant only for the native speaker group. The hierarchy matches the results that had been found in previous studies of complementizers, in that additional elements disfavour the use of the zero complementizer. Moreover, Italian speakers, the other group this factor was run for, also share the expected hierarchy.

Not only was the factor **additional elements in the matrix clause** found to be significant for the three groups where it was included in the variable rule analysis, but the hierarchy of factors was found to be the same in them as well. The distribution is what was predicted by Tagliamonte and Smith (2005); whereby elements in the matrix clause disfavour the use of the zero complementizer. Although this factor could not be examined for the French speakers, this was due to the fact that all of the tokens with additional elements in the matrix clause appeared with a *that* complementizer, so they too appear to share the same hierarchy.

**Tense** was not found to be significant for any of the groups, in terms of hierarchy, however, the English, French and Italian groups show the hierarchy (present>past) found by Tagliamonte and Smith (2005). This tendency is reversed in the German speakers where past forms have a higher factor weight than present forms. Note that the difference between these two factors is very small and that the lower overall number of tokens for the German speakers might have influenced this result.

## 7.8 Discussion

There are a few differences between the four groups in terms of the multivariate analysis results. Although lexical verb has the highest range in all groups, there is a switch in terms of the hierarchy of *think* and *hope* between the native and non-native groups. For the subject of the matrix clause, the native group shows a higher factor rate for second person subject than first person ones, while this tendency is reversed for the non-natives and matches what had been found in previous studies. Finally, for tense, the German speakers do not share the hierarchy of the other groups.

While these differences are important, the overall picture is one of similarity, however. The non-natives, for the most part, match native speaker patterns (either those of the e-mail control group or those considered in previous studies). *Think, tell, know* and *say* have higher rates of zero complementizer than other verbs, first person subjects are more likely to be used with a zero complementizer than other subjects, elements, either in the VP or the matrix clause, inhibit the use of zero. This demonstrates that the non-native speakers have acquired many of the constraints which operate on English complementizer patterns.

The (Swiss) German speakers have a zero form used in a similar way to the English zero complementizer form in their native language and they are the non-native group with the highest rates of zero complementizer. They are not the only group to show variability in the feature, however, as both the Italian and French groups also use the zero complementizer form. While the similarity between German and English might have benefited the German speakers in some ways, the fact that they and the other two non-native groups share native speakers' hierarchies of constraints and ranges is due to more than surface similarity. Although we do not know what constraints operate on German complementizers, it is unlikely that they are the same constraints as in English. The German speakers (as well as the French and Italian speakers) are applying English variable rules for their use of the zero complementizer form.

The rates of zero complementizer are lower across the board than in Tagliamonte and Smith's research. The main explanation for this is that Tagliamonte and Smith (2005) examined oral data while this analysis considered e-mails. This further underlines the extent to which e-mails, even relatively informal ones, cannot be considered on a par with speech or even written data especially in features where formality has been found to constrain the use of one of the variants.

In terms of the various cases of variability discussed in chapter 4, this feature is most likely to fall into case A, as for the most part the variable patterns of the native speakers

have been acquired by the non-native groups. This is very similar to what was found for relative pronouns and is especially interesting because these are two features in which the variability was not explicitly taught.

Case A	French, German and Italian present similar variation patterns to native speakers	variation acquired
--------	--	--------------------

## 7.9 Conclusion

This chapter has shown that despite differences in selection of the *that* and the zero complementizers forms not being explicitly taught to non-native Swiss speakers of English, these speakers have nevertheless acquired the variable rules of native speakers. The multivariate analyses of the various linguistic groups have shown that the hierarchies and ranges of the non-native groups are closely matched to those of the native control group. These patterns also match the constraints that have been uncovered in previous analyses of English zero complementizers (Tagliamonte and Smith, 2005).

Despite there being no similar zero complementizer form in two of the three source languages, the French, German and Italian speakers have been able to integrate the variable rules of English complementizer distribution. This underlines the fact that there are a number of underlying syntactic distribution patterns which non-native speakers can acquire without explicit teaching.

## 8. Lexical Variation II: Also, as well and too

### 8.1 Introduction

While the first case of lexical variation considered in this thesis (*information* vs. *informations*) was associated solely with interference-based variation, this chapter will discuss a case which is linked to target-based variation. The three additive adverbials *also*, *as well* and *too* (examples 1-3) are synonymous and Fjelkestam-Nilsson (1984) has shown that their distribution is conditioned by style and register as well as clause structure in native English.

- (1) maybe there's **also** some Porto wine. (c, Italian, e-mail)
- (2) I ordered **as well** some shiny days to make your stay in Zurich as comfortable as possible. (h, German, e-mail)
- (3) it was a situation that I shared, **too**. (f, French, e-mail)

The analysis in this chapter will examine the non-native speakers' use of these forms in order to establish whether they have rates of variation similar to native speakers. Because the variants are interchangeable and semantically equivalent there is a possibility that the non-native speakers will select just one of the variants and not use the other two. This is to some extent what has occurred for the future tenses; recall that *going to* is used far less often by the non-native speakers than by native speakers.

This feature is rather different from the other target-based variation features studied thus far because the difference between the variants is linked purely to lexis rather than to morphosyntax. The analysis will only focus on the occurrences of the three variants and does not consider clauses where one of the variants could have been used but was not (i.e. example 1').

- (1') maybe there's some Porto wine (modified from example 1)

The analysis of this feature will also allow us to consider aspects of interference-based variation in the non-native speakers as the variants are sometimes found in positions that are not acceptable to native speakers (examples 4-6).

- (4) he looked at **as well** the budget (h, German, e-mail)
- (5) She agrees, but she can't take **also** this job. (c, Italian, e-mail)
- (6) they have to go **also** for two months (u, French, e-mail)

## 8.2 Also, as well and too

The three variants *also*, *as well* and *too* are additive adverbials. Additive adverbials belong to the focussing class of adjuncts in Quirk, Greenbaum, Leech and Svartvik's categorisation of adverbials (1985:609). This category also includes *either*, *even*, *neither*, *nor*, *in addition*. I have decided to focus only on *also*, *as well* and *too* as they are the most semantically similar and thus function as lexical variants. Furthermore, Fjelkestam- Nilsson (1984) had focused primarily on these three variants in her research on additive adverbials in British and American English corpora.

A surface analysis of the Swiss data revealed that French, German and Italian speakers use the whole range of variants available, as the data revealed examples of *also*, *as well* and *too* in the output of the three linguistic groups (examples 7-15).

### Also

- (7) Actually, this message is **also** a big HELP request (j, French, e-mail)
- (8) 'c' **also** created some wonderful posters (p, German, e-mail)
- (9) The fifth of the IFMSA Committees is **also** existing in Switzerland (c, Italian, e-mail)

### As well

- (10) I talked again to 'X' today and she'd really like to visit those three cities **as well** (V, Italian, e-mail)
- (11) I think the grammar is changing **as well** now (S, French, interview)

- (12) I will try to get a video player **as well** for the meeting (h, German, e-mail)

### Too

- (13) it was a situation that I shared, **too**. (f, French, e-mail)  
 (14) we discussed that we could do it in CH , **too** and compare the results. (h, German, e-mail)  
 (15) and I hope it was interesting for you, **too**. (b, Italian, e-mail)

A first aim of the analysis, then, will be to establish whether the three variants are used at percentages similar to those of the native speakers. The analysis of the two future forms, *will* and *going to*, revealed that although both forms were used by the non-native speakers, *going to* was used far less frequently by the three non-native speaker groups than by native speakers. By comparing the rates at which *also*, *as well* and *too* are selected by the non-native and the native groups, this section will be able to uncover what the situation is in terms of these additive adverbials and whether, similarly to the future, one of the forms is selected instead of the other two variants.

Although the adverbials are semantically similar, there is a difference in terms of the positions where they are acceptable within the clause. While *also* can be used in a wide range of positions (Jacobson, 1964:226), the other two adverbials tend to be found only in clause final position (Jacobson, 1964:353) (table 8.1).<sup>114</sup>

Sentence	Positions
(Also) John (also) created some wonderful posters (also)	initial, mid, final
John (, too,) created some wonderful posters (too)	final
John (, as well,) created some wonderful posters (as well)	final

Table 8.1: Possible positions of *also*, *as well* and *too*<sup>115</sup>

<sup>114</sup> Jacobson presents his results in terms of the constituent referred to as well as position, but the positions in table 8.1 can be constructed from his figures.

<sup>115</sup> Even in the cases when *too* and *as well* seem to be placed in a more central position within the sentence, i.e. ‘c, *too*, created some wonderful posters’, they were still considered to be in final position and coded accordingly; this has to do with the fact that they are modifying only the NP in the clause and not the whole clause and, thus do indeed come after it.

Comparing the non-native placement of the variants to that of the natives will allow us to analyze the interference-based variability present in this feature. Moreover, in the case of *also* which can grammatically be placed in several positions, the percentages of placement distribution between the native and Swiss groups can be compared. This will allow us not only to establish how frequently the Swiss speakers use the variants in unacceptable positions, but also to see which variants and positions are most likely to be a source of difficulty.

### 8.3 Additive adverbials in the source languages

In terms of the number of additive adverbials synonymous to *also*, *as well* and *too* in the source languages, the situation is rather different, as French, German and Italian have only one main form equivalent to the three English adverbials.

That is not to say that the three non-native languages do not have synonyms for the main term, but that in each case there is one term which is the one predominantly used.<sup>116</sup> This is a quite different situation from English, where all three forms have been found to be used frequently both in speech and in writing (Fjelkestam-Nilsson, 1984).

The main form in French is *aussi*, with *également* used considerably less frequently; Italian has *anche* as the main form, with *pure* used alongside it (as well as the very literary form *altresi*); German has *auch* and a much less frequently used *ebenso* and Swiss German has *au* or *öi* depending on the dialect. Table 8.2 presents the differences between the Swiss languages and English schematically.

---

<sup>116</sup> The distribution of the various forms in the three Swiss languages has not in fact been studied; however, my own native speaker competence for French and discussions with native speakers of the other two languages has confirmed this view.

English	French	German	Italian
also			
as well	aussi	auch	anche
too			

Table 8.2: *Also*, *as well* and *too*: equivalents in the three Swiss languages

When selecting additive adverbials in English, the Swiss speakers will have three options while there is only one main form in their native languages. Consequently, it may be that they use one form more than the other two and more than native speakers would.

Another aspect linked to the English variants is the fact that *also* is a German word meaning *therefore*. The surface similarity between the English and the German terms will undoubtedly not be without consequence, but until the variability is studied in more detail we cannot say whether the similarity will inhibit its use or favour it.

Although the three variants are taught in Swiss schools, their distribution is not. Moreover, while the positioning of the three forms is different, this is never explicitly stated in English language teaching, despite the fact that Swiss students are expected to place the variants appropriately.

#### 8.4 Previous work on additive adverbials

Although to date, there is little research on non-native variation in the usage of the three adverbials in English, their distribution in native English has been studied quite extensively. Brita Fjelkestam-Nilsson's 1984 study is based on results obtained in three corpora which were first made available in the late 1970s. One of the corpora consists of



American written texts (Brown University Corpus, henceforth BUC, Kucera and Francis, 1967), one of British written texts (Lancaster-Oslo/Bergen Corpus, henceforth LOB, Johansson, 1978), and one of oral British data (London Lund Corpus, henceforth LLC, Svartvik and Quirk, 1980).

For her study, Fjelkestam-Nilsson compared the distribution of the three additive adverbials (only *also* and *as well* were considered for the oral data) in terms of text type and origin of the speakers. Her results will be very useful for the present study, even if there is a considerable time lapse between the two studies, as it will help determine whether the non-native groups are similar to native speakers.

The main results of Fjelkestam-Nilsson's research will be presented alongside those for the Swiss and native e-mail data, but note that the main differences that she found in the choice of *also* and *too* were stylistic: *also* was favoured in "informative" texts more than in "imaginative" ones, although it was used more often than *too* in all text types (Fjelkestam-Nilsson, 1984:16).<sup>117</sup> There were also some differences between American and British usage, and between writing and speech, but these were not as significant as the difference between types of text. She also found that the adverbial *as well* was at a mid point between the two other adverbials; less formal than *also* but not as informal as *too*.

## 8.5 Extraction and Coding

Tokens of *also*, *as well* and *too* were extracted from both the oral and e-mail components of the IFMSA corpus as well as from the native speaker e-mail corpus. Obviously, cases of *too* as an intensifier rather than as an additive adjunct were excluded from the analysis (example 16), similarly tokens of *also* where it functioned as a subjunct rather than an adjunct were excluded as well (example 17).<sup>118</sup>

(16) I cannot do *too* many things at the same time. (o, German, e-mail)

<sup>117</sup> The text types are those given in the corpora Fjelkestam-Nilsson used

<sup>118</sup> Almost all of the tokens of *also* as a subjunct were found in the native data.

- (17) *Also* don't forget about the town club on Friday evening at 7PM (native English e-mail)

The tokens were coded for the usual external factors of speaker, native language, register and corpus. In terms of internal factors, those considered were position within the sentence and whether the sentence was deemed non-native or not. This last factor group was perhaps somewhat subjective; for this reason, judgments were based on the intuitions of at least two native speakers.

## 8.6 Results

### Overall distribution

A total of 311 tokens of *also*, *as well* and *too* were extracted from the non-native corpus. Table 8.3, provides the overall distribution of the three variants considering oral and e-mail data separately.

Table 8.3: Non-native distribution of the three additive adverbials for source type (all groups combined)

	% of <i>also</i>	% of <i>as well</i>	% of <i>too</i>	Total N
E-mails	82	11	7	221
Oral Data	90	7	3	90
Total	84	10	6	311

A chi square test revealed that the differences between the oral and written segments of the corpus were not significant,<sup>119</sup> so they are combined in the rest of the analysis.

Table 8.4 considers the overall distribution of *also*, *as well* and *too* in terms of the three non-native language groups.

Table 8.4: Non-native distribution of the three additive adverbials by speaker group

	% of <i>also</i>	% of <i>as well</i>	% of <i>too</i>	Total N

<sup>119</sup> Chi-square = 3.26124765559715, *p* is less than or equal to 0.20. The distribution is not significant.

German	<b>51</b>	<b>40</b>	<b>9</b>	43
French	<b>81</b>	<b>6</b>	<b>13</b>	63
Italian	<b>92</b>	<b>4</b>	<b>4</b>	205 <sup>120</sup>
Total	<b>84</b>	<b>10</b>	<b>6</b>	311

In terms of the overall distribution, the Italian speaker tokens make up roughly two-thirds (205 out of 311) of the total. The number of tokens for the other two groups, though small, will allow us to gain some idea of their distribution of the variants. The main variant used by all three groups is *also*, although the German speakers do use it considerably less than the other two (at a rate of 51% opposed to 81% and 92% for the French and Italian speakers respectively). The French and Italian speakers have fairly low rates of the other two variants, whereas the German speakers use *as well* a full 40% of the time. The differences between the three groups are significantly different and the German speakers are not selecting the three variants in the same way as the other two groups.

The native e-mail corpus provided 47 tokens of the three adverbials. Table 8.5 considers the results for the native e-mail control group in order to establish whether the non-native patterns are typical of native speakers.

Table 8.5 Overall distribution of additive adverbials (native e-mail corpus)

	% of <i>also</i>	% of <i>as well</i>	% of <i>too</i>	Total N
E-mails	<b>68</b>	<b>17</b>	<b>15</b>	47

The results for the native e-mails show that similarly to the three non-native groups *also* is the form used most frequently. Table 8.6, which gives the figures from Fjelkстам-Nilsson's study for both oral and written data, will allow us to establish whether the e-mail results are comparable to previous findings.

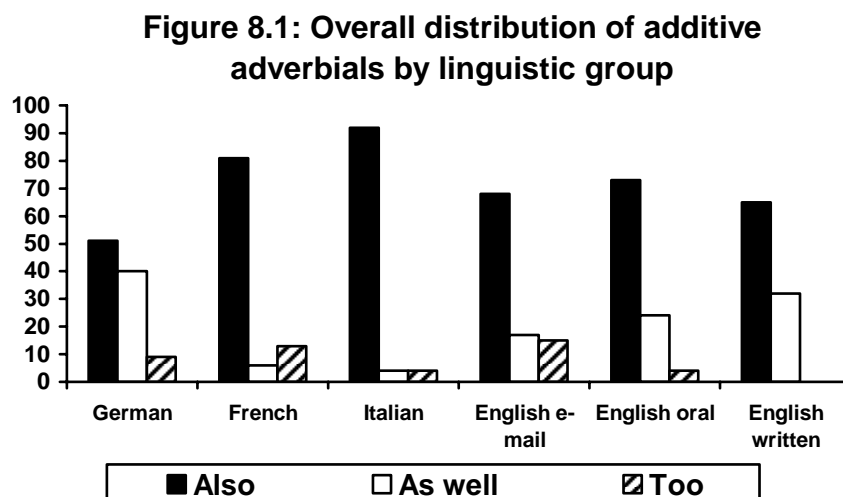
Table 8.6 Overall distribution of additive adverbials in Fjelkстам-Nilsson's results, 1984)

	% of <i>also</i>	% of <i>as well</i>	% of <i>too</i>	Total N
Written data (BrEng and AmEng)	<b>73</b>	<b>24</b>	<b>4</b>	2392

<sup>120</sup> The difference between the three language groups is **significant** with *p* being less than or equal to 0.001.

(LOB and BROWN)				
Oral data (LLC)	65	35	Not available	367

Although we do not have figures of *too* for the Fjelkstad-Nilsson's LLC data, the rest of the results look very similar to what had been uncovered for the native and non-native groups; namely that *also* is the main form. Figure 8.1 plots the results for the three non-native groups alongside the findings of the native e-mail corpus and of Fjelkstad-Nilsson's research (English oral and English written in figure 8.1).



The following points can be drawn from the first stage of analysis:

- *also* is the main variant for every single group and is selected more than 50% of the time, although the German speakers use it less than the other groups.
- the lower rate of *also* found in the German group may be linked to a lexically similar form in German
- *as well* is the next most favoured form for the native speakers and the German speakers
- *too* is the second most favoured form for the French speakers and the Italian speakers use *as well* and *too* equally

Although the French and Italian speakers do not use the *as well* and *too* variants as much as the native speakers, they use them frequently enough to establish that all three adverbials are part of their linguistic repertoire. We know this because they do not categorically use *also* but show some variation with the other two forms. This is very different from what was found for *will* and *going to* where the latter form was used less than 2% of the time.

The German speakers use *also* far less than the other groups and *as well* far more; this is most likely in part due to the presence of the lexical item *also* in German, but because Fjelkstad-Nilsson had found higher rates of *as well* in her oral corpus, it may be that the higher rate of *as well* in the German output reveals a closer similarity to native norms than the other two non-native groups.

In order to establish how similar the non-native groups are to native speakers, we need to consider the interference-based variability aspects of the feature. First we will need to determine how frequently the non-native speakers use the variants in a way that is not acceptable to native speakers and secondly, if there are non-native tokens, we need to uncover what is triggering them.

### **Interference-based variation**

An analysis of the non-native usage of the additive adverbials in the non-native corpora revealed two different types of problem involving *also*, *as well* and *too*. As well as tokens of the adverbials which were deemed to be positioned in an ungrammatical manner (examples 18-19), there were also cases of tokens where one of the variants was used with a meaning which would be considered unacceptable to a native speaker (example 20). Of the two types of problems, the one associated with problems of placement was more common than the second type.

(18) *Also* return date will be problem (p, German, e-mail)

(19) It is *as well* an occasion to thank 'a' another time (b, Italian, e-mail)

- (20) Another thing: I can't come to the meeting *also* on the 23-24 feb.<sup>121</sup> (c, Italian, e-mail)

Because the type which is associated with lexical choice was less common, it will be dealt with first, before focussing on the findings in terms of problems in position placement. The analysis found only 7 tokens of problems linked to lexical choice but two different types of it; 5 tokens where *also* was used in place of *either* or *neither* (example 20 above), and two tokens where *as well* is used in place of *as well as* (example 21). These tokens have been removed from the subsequent analysis. The low number of tokens of this type suggests that this aspect did not create many problems for the non-native speakers.

- (21) a link between the international world and the local world and *as well* to the Swiss medical students' association (b, Italian, e-mail)

In terms of the additive adverbials, there were problems in placement for only two of the variants. *Too* was always placed at the end of the clause so presented no difficulty for the non-native speakers, but *as well* was used by the Swiss speakers in other positions than final position. Specifically, 10 out of the total 29 tokens of *as well* for all three groups were placed in a mid position (as in example 19 above) and one was placed in initial position.<sup>122</sup> *Also*, which has a far greater range of acceptable positions, was nevertheless occasionally placed in such a way which made it sound unnatural to a native speaker, occurring either in an inappropriate medial position (example 22), or clause finally, in this case generally alongside another adverbial in the inappropriate order (example 23).

- (22) I hope that *also* the other folks out there will do the same (b, Italian, e-mail)

- (23) and ZH can take people *also* for two months (u, French, e-mail)

In all, 18% (55 out of 311) of the tokens were judged to be non-native in terms of positioning. Table 8.7 provides the breakdown of deviant tokens by speaker group.

<sup>121</sup> The context makes it clear that this speaker meant *either* and not *als*.

<sup>122</sup> Some native speakers of Canadian English do this as well.

Table 8.7: Problematic positioning of additive adverbials

	<i>Also</i> N	<i>As well</i> N	<i>Too</i> N	Percentage non-native	Total N (non-native)
Italian	34	2	0	18	36
French	8	0	0	13	8
German	3	8	0	26	11
Total	45	10	0	18	55

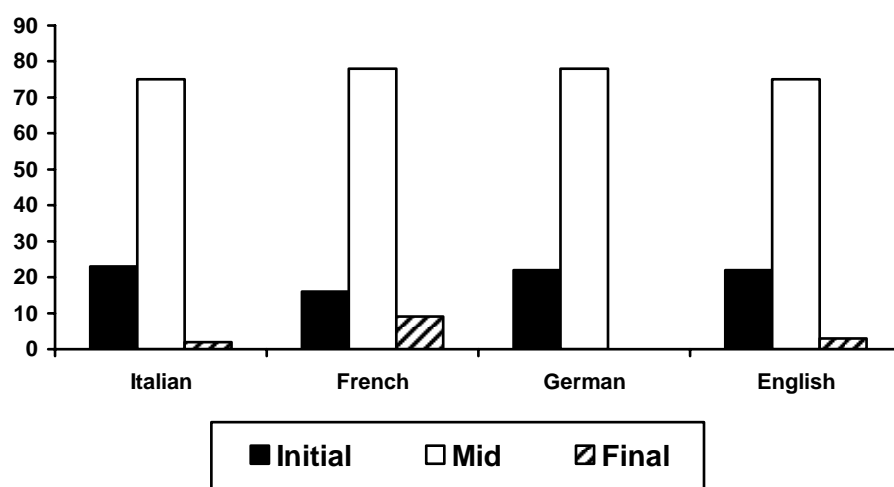
The overall rate of problematic tokens for the non-native speakers is fairly high, although the numbers are low. This demonstrates that the non-native groups are using the additive adverbials in positions unacceptable to native speakers despite using the three variants at rates similar to those found in English writing. The German speakers have the highest rate of non-native tokens at 26%. In terms of the specific variants placed in a non-grammatical fashion, *also* accounts for 80% of the non-native tokens. There are differences between the non-native groups, however, as the German speakers are far more likely to place *as well* in a position where it is not acceptable to native speaker. Not only are the non-native speakers using the variants differently to native speakers, but the German group does not share the patterns of the other two Swiss groups. This may be due to their overall higher rate of *as well*, by using it more frequently than the French and Italian speakers, there are more opportunities for them to use it in a problematic position.

### Position of *also*

As the only variant which can be found in several clause positions, *also* will now be considered on its own, in order to determine whether the non-native speakers follow the distribution patterns of native speakers. The positioning of adverbials is in large part linked to focus, so it may be that differences between groups will be tied to that; however, because the type of messages were similar across the groups, this will still give us an idea whether the groups share similar patterns. The ungrammatical tokens of *also* have been removed from consideration, so that only the tokens which have been judged native-like are compared to the native speakers' output. Table 8.8 and figure 8.2 provide the results for the different linguistic groups.

Table 8.8 Overall distribution of *also* by clause position

	% of Initial	% of Mid	% of Final	Total N
Italian	23	75	2	152
French	16	78	6	42
German	22	78	0	18
Non-native speakers combined	20	77	3	212
Native speakers (e-mail corpus)	22	75	3	32

Figure 8.2: Distribution of *also* by sentence position

The native and non-native rates are very similar in terms of positioning; the mid position is the most frequently selected position (around 75%), the initial position is the next most frequent position (around 20%) and the final position is very rarely used. There is no significant difference between the four groups,<sup>123</sup> so it looks to be the case that they share the same distribution. Although there are a number of tokens in ungrammatical position in the non-native data, the overall distribution of positions is very similar in the four groups.

<sup>123</sup> Chi-square = 2.22848503245692.  $p$  is less than or equal to 1. The distribution is not significant.



## 8. 7 Discussion

The results revealed a number of contradictory points so before discussing the reasons why this might be the case this section will briefly summarize the various findings.

- *also* is the main form used in all the linguistic groups
- however, *as well* is used far more frequently by the German speakers than the other groups.
- the non-native speakers all show high rates of interference-based variation as their positioning of the adverbials does not follow native norms.
- however, the French and the Italian speakers deviate from native norms mainly in their positioning of *also*, while the German speakers deviate in their positioning of *as well*.
- the three Swiss groups are very similar in terms of their positioning of *also*, with the mid position as the overwhelmingly favoured position.
- this matches what was found for the native e-mailers.

The wide range of results makes it extremely difficult to accurately determine what is happening. In some ways, this variable has not been acquired in a native-like manner by the Swiss groups, but at the same time there are also a number of factors which do appear to reveal similarities between them and the native group. Ultimately, the aspects which differentiate the non-native speakers from the native speakers outweigh those which are similar. For this reason, this variant is most likely a case of differentiated acquisition by the three groups - they have not acquired the patterns of the native speakers and furthermore there does not appear to be focussing between the groups (case C).

Case C	French, German and Italian vary unlike native speakers and unlike each other	variation influenced by native language only
--------	--	--

Although all three groups have similar problems, it would be misleading nevertheless to suppose that these similarities are due to much more than to the inherent difficulties of the features, rather than the fact that the speakers use English together and share the same

form of English. This is apparent by the fact that the German speakers show considerably different patterns to the other Swiss groups.

## 8. 8 Conclusion

The analysis of the three additive adverbials has highlighted the fact that a single feature may provide us with contradictory findings as to whether it has been acquired natively or not. Some aspects of a feature may be used in a native-like manner while other aspects remain very different. The non-native speakers do use all three forms, but not at the rates found in native speakers. Once the problematic sentences have been removed, the non-native speakers show that their positional distribution of *also* is very similar to native speakers. In terms of the additive adverbials, although one variant has not been unduly favoured over the others, as was the case for the future tenses, the native patterns cannot be said to have been fully acquired because there is still a high rate (18%) of tokens placed in a position not acceptable to native speakers.

How does this compare to the other lexically variable feature? Similarly to *information/s* the non-native speakers have been shown to use distinctive strategies to deal with the variability. The German speakers use more *as well* than the other two groups for example. Non-native speakers appear to have difficulty with lexical variation whether it is due to interference with their source language or variability directly in the target language.

## Part III

### **9. Discussion and conclusion**

#### **9.1 Introduction**

What have the analyses of the five variables considered in this dissertation shown us about how non-native speakers' deal with inherent variability? How have these analyses helped us determine whether focusing was occurring in the English of Swiss speakers? Furthermore, what overall impression can we obtain by considering all the variables at once?

This chapter will examine these questions to establish what we have uncovered about how non-native speakers deal with inherent variability and whether focussing is underway in Switzerland. The examination of the five different variables has provided us with several outcomes in terms of non-native variation and focussing. It is only when considering all the features analyzed in this thesis together that we will be able to draw some more general conclusions about the language variety/varieties under consideration.

In this chapter I will consider the findings of the various features studied in order to establish whether any general points emerge (section 9.2). I will also assess the implications of these results in terms of the two main aims of the thesis:

- to establish whether the English spoken in Switzerland has acquired a focused form (section 9.3)
- to discover how non-native speakers of English deal with inherent variability in the target language (section 9.4).

The final section of this chapter will discuss possible implications of the findings in the broader context of language variation and change as well as second language acquisition.

## 9.2 Summary of overall findings

In order to classify the English spoken in Switzerland, the three categories of variability which were presented in chapter 4 must be examined. Recall that features which were found to show the same variable rules, hierarchies and constraints in native and non-native groups were judged to have been acquired in a native-like manner (case A); those in which the non-native groups' variability was found to be different from native speakers, but in which the three Swiss groups could be shown to share the same variable rules, were judged to be examples of focussed variables (case B); and finally, features which were shown to have entirely different variation patterns in the four groups were judged to be variables in which the patterning was influenced solely by the native language (case C). Table 9.1 below repeated from chapter 4 presents this in schematic form.

Table 9.1: Potential cases of variation

Case	Patterns found	Outcome
Case A	French, German and Italian present similar variation patterns to native speakers	variation acquired
Case B	French, German and Italian vary unlike native speakers but like each other	focused form (or common problems of acquisition)
Case C	French, German and Italian vary unlike native speakers and unlike each other	variation influenced by native language only

Just as the individual features were judged to belong to one of the three categories, an attempt will now be made to classify the whole range of features considered together and find what can be said about the non-native varieties of English examined here in general.

Five distinct morphosyntactic and lexical variables were studied in this thesis and all three of the above mentioned cases were found in the results. The main findings for each feature are presented briefly below.

- Feature 1: **Non-count plural forms – *information* vs. *informations***

The use of *informations* in place of *information* was selected as a feature showing only interference-based variation. The results demonstrated that the three non-native groups had different patterns from one another, but also that, within the linguistic groups, the speakers themselves were very different, with some never using the non-native form and others using it almost exclusively. This was judged to be a case of variation influenced primarily by the native language (or more accurately by individual speakers) because there was no evidence of the variability being conditioned by any other factors.

Case C	French, German and Italian vary unlike native speakers and unlike each other	variation influenced by native language only
--------	--	--

- Feature 2: **Future tenses – *will* vs. *going to***

Although non-native speakers were taught that *will* is used for prediction and *going to* for the near future, they did not demonstrate patterns similar to those found in native English. For all three groups, *will* was selected almost categorically rather than *going to*. The near categorical use of *will* made it impossible to examine many of the factors which previous native-speaker studies had found to affect the variability.

However, the groups showed similar instances of interference-based variation (i.e. non-native usages of the forms); despite the low use of the *going to* variant, the analysis also revealed that 8% of the *will* tokens were used in a way that could be judged to be non-native. Moreover, the groups showed similar rates of contraction of *will* and *going to* to *'ll* and *gonna*. Because of the similarity between the non-native groups, this feature was judged to be a possible case of focussing.

Case B	French, German and Italian vary unlike native speakers but like each other	focussed form (or common problems of acquisition)
--------	--	---

- Feature 3: **Relative pronouns**

The analysis of relative pronouns provided two main findings. First of all, there were very few cases of interference-based variability in the data, under 2%. Secondly, variable rule analyses revealed that variability in the non-native groups was conditioned in the same way as the native group. Like native speakers, non-native relative pronoun choice was found to be linked to syntax, animacy, adjacency and definiteness. Moreover, the constraint ranking of the factor groups generally matched that of native speakers, suggesting that it is likely that the non-native speakers shared the same rules as the native speakers, making it an example of a feature where the variation had been acquired in a native-like manner.

Case A	French, German and Italian present similar variation patterns to native speakers	variation acquired
--------	--	--------------------

- Feature 4: **Complementizers**

Like the relative pronouns, complementizers were revealed to be an area where the non-native speakers also very closely matched the variable rules of the native speakers. The non-native speaker patterns were found to be linked to lexical verb, grammatical person, tense, and the presence of additional elements in the verb phrase and matrix clause. Again, their constraint rankings and ranges corresponded to those of native speakers for the most part. The similarity between the native and non-native speakers was somewhat surprising given that both variants are acceptable in all situations and that the variability seems not to have been overtly taught. This feature was judged to be another case where non-native speakers had been able to acquire inherent variability in the target language.

Case A	French, German and Italian present similar variation patterns to native speakers	variation acquired
--------	--	--------------------

- Feature 5: *Also, as well* and *too*

The analysis of the three additive adverbials showed that the non-native speakers had a number of patterns in common with the native group; *also* was the form selected most often in all groups, for example. However, this feature showed a high rate of interference-

based variation as well: the non-native speakers were not always able to position *also* (and *as well* for the German speakers) in a native-like manner. Overall, the differences between the non-native speakers and the native group and between the groups themselves were such that the feature was judged to be involved in variation affected by the native language only.

Case C	French, German and Italian vary unlike native speakers and unlike each other	variation influenced by native language only
--------	--	--

As can be seen from this summary, three cases are represented across the variables: both the non-count plural forms and the additive adverbials fall into case C, the future falls into case B, whilst relative pronouns and complementizers fall into case A.

Considering some of the reasons why the different variables pattern differently may help us gain some understanding of the processes currently occurring in the English spoken in Switzerland.

### Feature type

Firstly, we can note that the two features of lexical variation (*informations* and the additive adverbials, *also*, *as well* and *too*) are those which show the most distinct patterns between linguistic groups. Neither of these features was used in a native-like manner by the non-native groups and the three Swiss groups were very different from each other. Secondly, the features in which the variants were syntactically conditioned - the relative pronouns and the complementizer forms - showed patterns very close to those of native speakers in all three non-native groups. Thirdly, future tense use, which is constrained both lexically and syntactically, appears to be intermediate between the other two types, in that it was not acquired natively but was found to be very similar across non-native speaker groups. Note that while *will* and *going to* share a similar meaning, there are nevertheless both syntactic and semantic factors that condition their use. We can conclude that, of the features considered in this thesis, those associated with lexical variation appear to have been more difficult for non-native speakers to acquire than those

in which the variation is tied to morphosyntax.

### Frequency

One reason for this might have to do with frequency. Lexical features, more specifically those that are restricted to single lexical items, such as was the case of the two in this thesis, provide considerably lower numbers of token than the morphosyntactic features and may therefore be more difficult to acquire natively. This point can be illustrated from our current data. Table 9.2 presents the numbers of tokens per feature<sup>124</sup> as well as the category they fall into and the percentage of tokens judged non-native (averaged across the three Swiss linguistic groups).

Table 9.2: Individual features by category, number of tokens and % of non-native tokens

Feature	Category	Total Number of tokens (all non-natives combined)	Average percentage of IBV
Non-count nouns (information/s)	Case C	119	43%
Future	Case B	1166	8%
Relative Clauses	Case A	547	2%
Complementizers	Case A	576	0%
Also/as well/too	Case C	221	20%

The two lowest frequency features are precisely the two which are the least well acquired in that they show neither native-like patterns nor cross-linguistic similarities. More frequent constructions are more likely to show native-like patterns of variability because the non-native speakers are also more likely to hear them and use them (Ellis, 2002, Tarone, 2002). The future variable, which occurs the most frequently, does not quite fit this pattern however. Despite appearing more than 10 times more often than *information* tokens and twice as often as relative clauses and complementizer forms, the non-native speakers do not use the future variants in the same way as native-speakers. It is possible that very high rates of occurrence are conducive to focussing rather than native-like acquisition; however, as the future is the only feature found to fall in category B, it is

<sup>124</sup> NB: in order to make these figures as comparable as possible the number of tokens only comprises those found in the e-mail data as the oral data was not considered in for relative clauses and complementizer analyses.



difficult to assert this with certainty. Lexical items which are only seldom encountered will not provide the same opportunities both in terms of matching native speaker patterns and focussing with the other linguistic groups.

### **Interference-based aspects**

Table 9.2 also shows that there is some correlation between the number of interference-based tokens and their outcome in terms of type of variability. The features with the highest amounts of interference-based variability (between 20 and 43% here) are also the features which show the least similarity to the native speaker patterns, while the features with the lowest amounts (0 to 2% in the present data) are the most likely to be found to have acquired the native variable rules. To some extent this is to be expected, as features non-native speakers have difficulty with will not be acquired in a native-like manner.

However, as shown in the analysis of *will* and *going to*, these differences between native and non-native groups persist even when the interference-based tokens are removed from consideration. Recall from chapter 5 that the non-native rates of *will* usage were far higher than native rates even after the tokens of *will* overuse were removed.

At this stage, it is not apparent which of these two factors is the cause and which is the effect; do we find high rates of IBV because the feature has not been acquired natively or do the high rates of IBV inhibit a native-like acquisition of the feature? This undoubtedly merits further research as it would help us understand some of the sources of difficulty for second language learners.

In terms of the features considered in this thesis, the factors of frequency and amount of interference-based variation appear to be linked; the more frequently a feature is used, the less likely it is to be used in a non-native fashion and the more likely the non-native speakers are to be able to acquire the native patterns governing its use. Again, this cannot be disentangled from feature type, as the features associated with lexical variation had more interference-based variation than the morphosyntactic features.

### 9.3 Implications of the results for aim 1: focussing in Switzerland

Although the results have demonstrated that several varieties can be shown to share patterns, there is no sign of widespread focussing taking place in Switzerland. Indeed, the results argue fairly conclusively against much current focussing in Switzerland: of the features that were shown not to have acquired native-like patterns, only one was found to have been a potential case of focussing, while the other two features showed considerable differences between the groups. For the features of relative pronoun selection and complementizer deletion, the Swiss groups are similar to each other, but the fact that they also match the native speakers means that the extent to which they contributed to each others' native-like acquisition cannot be determined.

Of course, the fact that none of the features considered in this thesis have shown that the English spoken in Switzerland is particularly focussed does not necessarily mean that other features would not have provided a different conclusion. It is also possible that other groups of Swiss English speakers might have provided us with different results. In terms of the standard of English spoken and the frequency at which it was used, IFMSA-CH is fairly representative of the overall Swiss situation, however. If an association where English is used everyday (albeit through the medium of e-mails) has not reached a stage where a focussed variety is used, then this is not too likely to be the case.

Chapter 2 discussed some of the reasons why Switzerland could be considered to be similar to countries, such as India and Nigeria, where focussing has already taken place. Given that our results show that Switzerland is not very involved in focussing, despite similarities with these countries, we now need to consider aspects which differentiate it from these countries.

#### **Time**

A first reason is linked to the length of time that English has been in use as a lingua franca in Switzerland. The fact that we were not able to conclusively show that the varieties of English used in Switzerland share to any great extent specific norms and patterns that differentiate it from native English does not mean that the English in

Switzerland will not at some point acquire focussed forms and indeed come to be considered a form of English in its own right. Indeed, many studies of new dialect formation (for example, Trudgill's research into New Zealand English, 2004) reveal that focussing takes places over several generations. Hence, the language of the first generation of native speakers of a new linguistic variety is not the same as that of the second and subsequent generations. It is more volatile and much more prone to showing idiosyncrasies and it is only in the language of the ensuing generations of speakers that we can hope to discover which features have been taken up in the new dialect.

In terms of Outer Circle varieties of English, the timeline for the formation of a new variety appears to be even longer; South Asia's first regular contact with English was in the seventeenth century and during the period of British sovereignty from 1765 until India's independence in 1947 English was used in administration and education (Crystal, 1995:101). English was used in Nigeria and other West African countries from the late nineteenth century onwards. English has been in use in both these countries for more than a hundred years, considerably longer than the three generations found in studies of new dialect formation and even more so compared to the length of time English has been in use in Switzerland. All of the varieties of English mentioned as belonging in the Outer Circle have long traditions of English as a second language, so these varieties of English have had time to change and focus over subsequent generations. On the other hand, we have found that the importance of English in Switzerland dates only from after the Second World War. It may then be that we were unable to posit a clear case of focussing, not because there is no focussing underway in Switzerland, but because it is too soon to be able to see the focussing. The fact that we found one example of apparent focussing suggests that this might indeed be the case.

### **English as a colonial language**

The manner in which English was first introduced to a number of Outer Circle countries may also help understand how new varieties come into being. These countries were for the most part British colonies and English played an important part in their government. In most cases, there was also an influx of English native speakers into these countries.

This meant that non-native speakers of English in the former colonies tended to use English directly with native speakers in a first instance, which is somewhat different from what is found in Switzerland. Although Swiss speakers occasionally use English with native speakers, most of the time they use it with other non-native speakers (Durmüller 2002).

As a result, although English was taught in schools in countries such as India and Nigeria, it was also learnt in a less structured way through direct contact with native speakers, which is quite different from Switzerland where English is acquired almost entirely through schooling. We can also assume that in the former colonies the first generations of non-native English speakers used English to a greater extent and more often.

### **Domains**

Another aspect which separates Switzerland from countries which have new focussed varieties of English has to do with the domains in which English is used. Although English is used every day in some Swiss companies, people who share the same language would only rarely use English with each other in Switzerland, where English is still primarily a lingua franca and not a language of everyday communication. Swiss people who use English speak it or write in it for work and as a tool of wider communication, but for most it is not a language used in the family or with friends. This contrasts with countries such as India and Nigeria. Not only did a large number of expatriates extend the use of English in the former colonies (Crystal, 1995:103), but government business and basic schooling were conducted in English. As noted by Kachru and Nelson (1996:81) 'in India or Singapore, use of English ranges from personal domains [...] to business, education, administration, creative writing and journalism. The result is that English has social penetration, that is, depth.' They further note that 'in Nigeria, the situation is essentially the same' (Kachru and Nelson, 1996:81). Of the domains listed by Kachru and Nelson, the business domain is the only one that can clearly be applied to the situation of English in Switzerland.

It is reasonable to conclude then, that focussing can only occur when a language is used in a wide range of contexts, or domains, and has acquired depth. Although English is important in Switzerland, it has not reached a point where it can be said to have upstaged the official languages.

With time and as the importance of English in Switzerland grows then it may be the case that Swiss speakers will begin to use a focussed form of English. As such the analyses presented here and the data collected for the project as a whole are an invaluable tool for future research; if the English in Switzerland reaches a focussed stage, it will come from groups of speakers such as the ones considered here and the results presented in this thesis will provide information into the incipient stages of the focussing of Swiss English. The data collected for this thesis will be to provide comparative data for the future.

### **English in Europe**

These findings are important also in terms of research into English as a lingua franca in Europe (Seidlhofer, 2001, 2005, 2006, Erling, 2004); if Switzerland has not yet reached a stage where focussing can occur, then other non-native varieties of English found in Europe must be considered to be at an even earlier stage of the process. The Swiss situation can then provide some insight into the direction where the use of English in the whole of Europe is headed. Many of the features initially considered by the project to be potential cases of Swiss English, such as *information* used as a count noun and the use of demonstrative *this* with both singular and plural nouns of have been found in VOICE – the Vienna Oxford International Corpus of English (Seidlhofer, 2005:92). The project's research into English in Switzerland can help researchers of English as a lingua franca add more to their understanding of how and why non-native speakers use English in Europe today.

The results of this thesis also can offer a slightly different perspective of non-native English use which can help inform other researchers of English in Europe, namely that of how non-native speakers deal with the inherent variability found in English. Much of the research on English in Europe (most importantly Barbara Seidlhofer's work on VOICE –

the Vienna Oxford International Corpus of English) has focussed not only on collecting empirical data of English as a lingua franca, but also on identifying ‘lexicogrammatical features of ELF,’ noting that ‘various features which thus far have simply been regarded as ‘non-standard errors’ are produced regularly by ELF speakers from many different first language backgrounds without posing an obstacle to communicative success’ (Seidlhofer, 2006). As this thesis has found, differences between native and non-native English speakers can be found at deeper linguistic levels as well, so by considering this type of variability researchers in Europe may find other signs of a distinctly European variety of (non-native) English.

#### **9.4 Implications of the results for aim 2: Second Language Variation**

The results of this thesis have shown that features were not all acquired in a native-like manner, even in cases where the variability had been overtly taught. In fact, the two features in which much of the inherent variability was not presented to non-native English students (i.e. complementizers and relative pronouns) were demonstrated to most closely match native patterns. On the other hand, the future was not acquired natively, despite the fact that both variants had been presented in class. It is difficult to assess why this might have been the case, but it is possible that the overt teaching of when one variant should be used and when the other should might have hindered the unconscious acquisition of the variable patterns rather than helping. The finding that variability can be acquired by non-native speakers in some cases is rather different from what had been uncovered in some of the earlier studies of Second Language Variation. Non-native speakers of French were found to have considerable difficulty acquiring native variable patterns (Rehner, Mougeon and Nadasdi, 2003, Regan, 1995), most specifically in terms of when and how frequently to use the more vernacular variants but also in replicating native speakers’ variable rules for features tied to inherent variability (Mougeon, Rehner and Nadasdi, 2004:421).

The crucial difference between these earlier studies and the present thesis is the type of native variability selected. Most of the features in the earlier studies were of variables in

which one of the variants was more informal (for example the rates of *ne* deletion by students of French in Regan 1995) and the studies found the non-native speakers did not match the native rates in informal contexts. The features selected in the present thesis, however, tend to be variable primarily in terms of internal, not external factors. Consider, for example, complementizer use; the main factors conditioning variability were factors such as lexical verbs and grammatical person, not gender or age. The results suggest that internal linguistic constraints may be easier to acquire than external, social constraints. It is important to remember that the French non-native speakers in Mougeon, Rehner and Nadasdi (2004) did not match native French Canadian speakers even if features which were not highly stylistically constrained, however.

Another possible difference could be linked to the proficiency of the speakers. Although most Second Language Variation research focuses on high-level non-native speakers, for the most part they are still language learners. The English speakers in the present thesis were no longer studying English but simply using it and are already high level bilinguals.

Considering the variation patterns of different non-native linguistic groups at the same time also allowed for a greater understanding of how non-native speakers deal with native variability. It allowed us establish the relative effect of the source language on English variability in a way that is not possible when examining a single non-native group. For example, it showed that native-like patterns were not necessarily linked to similar structures in the source language: native complementizer patterns are found in French, German and Italian speakers, although only German speakers have a zero complementizer form comparable to English. The non-native speakers follow the English patterns, not because they have similar ones in their own languages but because they are able to integrate the patterns of English.

The results of this thesis suggest that studying how non-native speakers deal with variability inherent in the target language can help us understand the processes of second language acquisition as it can help us understand which features are likely to be acquired with ease and which ones are not. Considering features that are variable in the target

language can provide researchers with an understanding of non-native language acquisition from a radically different perspective from the knowledge that can be gained from considering only interference-based features. For example, not only have we uncovered that non-native speakers make relatively few errors in selecting relative pronouns, but we now know that they are also able to acquire the less tangible aspects of relative pronoun selection, which can tell us how non-native speakers acquire deeper structures within language. This is especially true when considering how the rates of interference-based variability appear to affect how non-native speakers deal with target-based variation aspects as well. Non-native speakers are sensitive to the unconscious patterns that native speakers have and the analyses in this thesis have shown that for some features they are able to model them.

Another important finding is that, although there are cases where the non-native speakers have not acquired a feature in a native-like manner, there is far less directly ascribable interference than might have been expected. Although the structures of the source languages had at least some part in the ease or difficulty of acquisition, there is not much evidence of specific constructions being transferred from the source language. For example, the fact that French has relatively similar future constructions to English (in terms of the *going to* variant at least) did not help the French speakers use the comparable English variant, in fact they did not use *going to* at the same rate as the native speakers. Although the *going to* form is more frequent in conversational Swiss French, as it is in English, this did not affect the results, as the French speakers' rate of *going to* did not increase in the interviews or the meeting.

## 9.5 Additional Considerations

As well as the points discussed above, there is one further aspect in which this thesis contributes to our understanding of language; namely in terms of its use and analysis of e-mail communication. It provides one of the first (if not the first) in-depth analyses of the language used by an online speech community (or speech communities if the native e-mail data is considered as well). Moreover, this analysis looks at how non-native



speakers can use English as a means of constructing virtual identities and virtual communities.

In terms of the debate of where to place computer-mediated communication on the stylistic continuum of speech and writing, this thesis has found that, although online data shows higher rates of stylistically formal features than unmonitored speech, the rates are nonetheless not directly comparable to written data. E-mail communication is distinct from both speech and writing and further studies will be able to determine precisely how this type of virtual communication should be classified stylistically. Many of the variable rules which are found in speech have been found to be replicated in the e-mail data, however, and this shows there are still points of contact.

With the increase in the number of online communities and the increase of computer-mediated communication, there will be an growing need to examine how language is used online. Not only in terms of variability as has been focussed on here but also in terms of how language can be used to construct online identities and how language can be used almost exclusively for online purposes.

## **9.6 Closing words**

Although the analyses presented in this thesis showed that the English spoken in Switzerland is not focussed, they nevertheless have revealed how Swiss speakers use English and how non-native speakers, in general, can acquire different aspects of variability. The English spoken in Switzerland is not obviously involved in the process of focussing yet and at many points each linguistic group is still affected by their native language; however, the analyses have shown that the contact between the groups has influenced their English (in terms of the future tense) and, with time, may well affect it even further, as has been shown in other non-native English varieties.

This thesis has also served to further demonstrate the necessity of examining non-native speaker variability, not only in terms of the interference of the source language but also in

terms of the variability found in the target language. If we consider only those features which do not match native norms, we neglect a range of features which reveal much about language processes. Further research into second language variation will allow researchers to uncover which features can be relatively easily acquired by non-native speakers, in addition to our findings that internal factors are more likely to be matched by non-native speakers than external, stylistic factors, and might also shed light on how native speakers learn variation.

Non-native speakers are capable of dealing with some the inherently variable features of English; irrespective of whether the variability was taught, the non-native groups of Switzerland have shown that they are aware of and able to match the variable rules of native speakers in complementizer deletion and relative pronoun selection. A greater understanding of which features are most likely to be used natively, and why, will further our understanding of how variation is used by non-native speakers.

## References

- Bailey, G. (2002). Real and apparent time. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, (pp. 312-332). Oxford: Blackwell.
- Bailey, G., Maynor, N. and Cukor-Avila, P. (eds.) (1991). *The Emergence of Black English: Text and Commentary*. Amsterdam: John Benjamins.
- Ball, C. (1996). A diachronic study of relative markers in spoken and written English. *Language Variation and Change* 8:227-258.
- Ball, R. (2000). *Colloquial French Grammar: A Practical Guide*. Oxford: Blackwell.
- Baugh, J. (1983). *Black street speech: Its history, structure, and survival*. Austin: University of Texas Press.
- Bayley, R. (1995). Competing constraints on variation in the speech of adult Chinese learners of English. In R. Bayley & D. R. Preston (eds.), *Second Language Acquisition and Linguistic Variation* (pp. 97-120). Amsterdam: John Benjamins.
- Bayley R. (1999). Relativization Strategies in Mexican-English. *American Speech* 74(2): 115-139.
- Bayley, R. (2002). Quantitative Paradigm. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, (pp. 117-141). Oxford: Blackwell.
- Bayley R. and Preston D. R. (eds.) (1995). *Second Language Acquisition and linguistic variation*. Amsterdam: John Benjamins.
- Bayley, R. and Regan V. (2004). Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics* 8(3):323-338.
- Berglund, Y. (1997). Future in Present-day English: Corpus-based evidence on the rivalry of expressions. *ICAME Journal* 21:7-20.
- Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge: Polity Press.
- Busslinger L. (2004). La Suisse alémanique favorise l'anglais à l'école au détriment du français. 2004: le choix des langues. *Le Temps* (10 January 2004).
- Bürgi, H. and Bader, C. (2001). *Swiss Catalogue of Learning Objectives for Undergraduate Medical Training*. Retrieved 24 May 2006 from: <http://www.sgam.ch/pdf/Lernzielkatalog.pdf>

- Bybee, J., Perkins, R. and Pagliuca, W. (1991). Back to the future. In B. Heine & E. Traugott (eds.), *Approaches to grammaticalization*. (pp. 17-58). Amsterdam: John Benjamins.
- Cenoz J. and Jessner, U. (eds.) (2000). *English in Europe: The Acquisition of a Third Language*. Clevedon: Multilingual Matters.
- Chambers, J. K. (1995). *Sociolinguistic Theory*. Oxford: Blackwell.
- Chambers, J. K. (2002). Studying Language Variation: An Informal Epistemology. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, (pp. 3-14). Oxford: Blackwell.
- Chambers, J. K., Trudgill, P. & Schilling-Estes, N. (Eds). (2002). *The Handbook of Language Variation and Change*. Malden and Oxford: Blackwell Publishers.
- Cheshire, J. (ed.) (1991). *English around the World: Sociolinguistic Perspectives*. Cambridge University Press.
- Cheshire, J. and Moser L.-M. (1994). English and symbolic meaning: The case of advertisements in French-speaking Switzerland. *Journal of Multilingual and Multicultural Development* 17:451-469.
- Cossy, C. (2005). A Zoug, l'opposition au français au primaire est née de l'amour pour les travaux manuels. *Le Temps* (12 December 2005).
- Coupland, N. (1984). Accomodation at work: Some phonological data and their implications. *International Journal of the Sociology of Language* 46:49-70.
- Crystal, D. (1995). *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.
- Crystal, D. (2003). *English as a Global Language* (Second Edition). Cambridge: Cambridge University Press.
- Dewaele, J-M. (2004). Retention or omission of the *ne* in advanced French interlanguage: The variable effect of extralinguistic factors. *Journal of Sociolinguistics* 8(3):433-450.
- Dewaele, J-M. and Regan V. (2002). Maîtriser la norme sociolinguistique en interlangue française: le case de l'omission variable de 'ne'. *French Language Studies* 12:123-148.

- Dictionnaire Historique de la Suisse [Swiss Historical Dictionary]. (2003). Répartition des langues 1990 [Language distribution 1990]. Retrieved 4 August 2003 from: <http://www.sn�.ch/dhs/externe/protect/francais.html>
- Dillard, J. (1972). *Black English: its history and usage in the United States*. New York: Random House.
- Dürmüller, U. (2001). The presence of English at Swiss universities. In U. Ammon (Ed.), *The Effects of the dominance of English as a language of science on the non-English language communities*. Berlin-New York: Mouton de Gruyter.
- Dürmüller, U. (2002). English in Switzerland: From Foreign Language to Lingua Franca. In D. Allerton, P. Skandera, C. Tschichold (Eds.), *Perspectives on English as a World Language* (pp. 114-123). Basel: Schwabe.
- Eckert, P. (2000) *Linguistic Variation as Social Practice*. Oxford: Blackwell.
- Edwards, J. (1995). *Multilingualism*. London: Penguin.
- Erling, E. (2004). Globalization, English and the German University Classroom: A Sociolinguistic Profile of Students of English at the Freie Universität Berlin. Unpublished PhD thesis, University of Edinburgh.
- Eisenstein, M. (1989). *Dynamic Interlanguage: Empirical Studies in Second Language Variation*. New York: Plenum.
- Ellegard, A. (1978). *The Syntactic Structure of English Texts: A Computer-based Study of Four Kinds of Text in the Brown University Corpus*, Goteborg, Sweden: Acta Universitatis Gothoburgensis.
- Ellis, N. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24:143-188.
- Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Elsness, J. (1984). That or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies* 65: 519-533.
- Emirkanian, L. and D. Sankoff. (1985). Futur simple et le futur périphrastique. In M. Lemieux and H. Cedargren (eds.), *Les tendances dynamiques du français parlé à Montréal*. (pp. 189-201). Québec: Gouvernement du Québec.

- Ferguson, C. (1959). Diglossia. *Word*, 15, 325-340.
- Fjelkestam-Nilsson, B. (1983). *Also and Too: A Corpus-Based Study of Their Frequency and Use in Modern English*. Stockholm Studies in English, Stockholm: Almqvist and Wiksell International.
- Flanigan, B. & Inal, E. (1996). Object relative pronoun use in native and non-native speakers of English: A variable rule analysis. *Language Variation and Change* 8:239-262.
- Fox, A. (1990). *The Structure of German*. Oxford: Oxford University Press.
- Giles, H. (1973). Accent Mobility: A Model and Some Data. *Anthropological Linguistics*, 15: 87-105.
- Godfrey, E. and Tagliamonte, S (1999). Another piece for the verbal-s story: Evidence from Devon in southwest England. *Language Variation and Change* 11(1):87-121.
- Görlach, M. (1998) *Even more Englishes. Studies 1996-1997*. Amsterdam: John Benjamins Publishing Company.
- Görlach, M. (2002). *Still more Englishes*. Amsterdam: John Benjamins Publishing Company.
- Guy, G. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov, ed., *Locating language in time and space* (pp. 1-36). New York: Academic Press.
- Guy, G. (1988). Advanced varbrul analysis. In K Ferrara, B. Brown, K. Walters and J. Baugh (eds.) *Linguistic Change and contact* (pp.124-36). Austin, Texas: Department of Linguistics, University of Texas at Austin.
- Guy, G. & Bayley R. (1995). On the choice of relative pronouns in English. *American Speech* 70:148-162.
- Herring, C. (2001). Computer Mediated Discourse Analysis. In D. Schriffrin, D. Tannen and H. Hamilton (eds.), *The Handbook of Discourse Analysis* (pp. 612-634). Oxford: Blackwell.
- Hoffman, C. (2000). The spread of English and the growth of multilingualism with English in Europe, in J. Cenoz and U. Jessner (eds.) *English in Europe: the Acquisition of a Third Language* (pp. 1-21). Clevedon: Multilingual Matters.

- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (eds.), *Sociolinguistics*. Harmondsworth, England: Penguin.
- International federation of medical Students' associations. (2003). Retrieved 24 March 2003 from: <http://www.ifmsa.org>
- IFMSA-CH. (2003). Retrieved 24 March 2003 from: <http://www.ifmsa-ch.org>
- Jacobson, S. (1964). *Adverbial Positions in English*. Stockholm, Sweden: AB Studentbok.
- Jacobson, S. (1975). *Factors Influencing the Placement of English Adverbials in Relation to Auxiliaries: A Study in Variation*. Stockholm, Sweden: Almquist and Wiskell International.
- Johansson, S. (1978). *Manual of Information to accompany the Lancaster/Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Department of English, University of Oslo.
- Kachru, B. (ed.) (1982). *The Other Tongue*. Pergamon Institute of English.
- Kachru, B., and Nelson, C. (1996). World Englishes. In S. McKay and N. Hornberger (eds.), *Sociolinguistics and Language Teaching* (pp. 71-102). Cambridge: Cambridge University Press.
- Karstadt, A. (1996). Relative markers in Swedish-American English: Evidence for a contact language phenomenon? *American Speech* 71:27-48.
- Kerswill, P. (2002). Koineization and Accommodation. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, (pp. 669-702). Oxford: Blackwell.
- Kerswill, P. and Williams, A. (2000). Creating a new town koine: children and language change in Milton Keynes. *Language in Society* 29:65-111.
- Kucera H. and Francis, W. (1967). *Computational Analysis of Present-Day American English*. Providence: Brown University Press.
- Labov, W. (1966). *The Social Stratification of English in New York City*. Washington DC: Center for Applied Linguistics.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1994). *Principles of Linguistic Change. Volume 1: Internal Factors*. Oxford: Blackwell.

- Labov, W. (2001). *Principles of Linguistic Change. Volume II: Social Factors*. Oxford: Blackwell.
- Le Page, R. B. & Tabouret-Keller, A. (1985). *Acts of Identity*. Cambridge: Cambridge University Press.
- Lyster, R. (1996). Question forms, conditionals, and second-person pronouns used by adolescent native speakers across two levels of formality in written and spoken French. *The Modern Language Journal* 80: 165-182.
- Meyerhoff, M. (2002). Communities of Practice. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, (pp. 526-548). Oxford: Blackwell.
- McArthur, T. (1998). *The English Languages*. Cambridge University Press.
- Milroy, J. and Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21, 339-84.
- Milroy, L. (1987). *Observing and analysing natural language*. Oxford: Blackwell.
- Milroy, L. (2002). Social Networks. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, (pp. 549-572). Oxford: Blackwell.
- Mougeon, R. & Rehner, K. (2001). Variation in the spoken French of Ontario French immersion students: The case of *juste* vs *seulement* vs *rien que*. *Modern Language Journal*, 85:398-414.
- Mougeon, R., Rehner, K and Nadasdi, T (2004). The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics* 8(3):408-432.
- Nagy, N., Blondeau, H. & Auger, J. (2003). Second language acquisition and "real" French: An investigation of subject doubling in the French of Montreal Anglophones. *Language Variation and Change* 15(1):73-103.
- OFS (Office Federal de la Statistique - Swiss federal statistics office). (2002). Languages 2000. Retrieved March 3, 2002 from:  
[http://www.statistik.admin.ch/stat\\_ch/ber16/eufr16.htm](http://www.statistik.admin.ch/stat_ch/ber16/eufr16.htm)
- OFS (Office Federal de la Statistique - Swiss federal statistics office). (2005). Les Langues en Suisse: Vers la fin du Quadrilinguisme. Retrieved June 1, 2005 from:  
<http://www.bfs.admin.ch/bfs/portal/fr/index/themen/volkszaehlung/uebersicht/blank/medienmitteilungen.html>



- Penner, Z. and Bader, T. (1995). Issues in the Syntax of Subordination: A Comparative Study of the Complementizer System in Germanic, Romance, and Semitic Languages with Special Reference to Bernese Swiss German. In Z. Penner (ed.), *Topics in Swiss German Syntax* (pp.73-290). Bern: Peter Lang.
- Phillipson, R. (2003). *English-only Europe: Challenging language policy*. London: Routledge.
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics*, 18, 581-618.
- Poplack, S. (2000). *English History of African American English*. Oxford: Blackwell.
- Poplack, S. and Dion, N. (2004). The French *Future* in Grammar, Thought and Speech. Paper presented at New Ways of Analyzing Variation Conference [NWAV 33], Ann Arbor: University of Michigan.
- Poplack S. and Tagliamonte S. (1999). "The grammaticization of *going to* in (African American) English", *Language Variation and Change*, 11(3):315-342.
- Poplack, S. and Tagliamonte, S. (2001). *African American English in the Diaspora*. Oxford:Blackwell.
- Quirk, R. (1957). Relative clauses in educated spoken English. *English Studies* 38:97-109.
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman, Harcourt.
- Rash, F. (1998) *The German Language in Switzerland: Multilingualism, Diglossia and Variation*. German Linguistics and Cultural Studies Volume 3. Bern: Peter Lang.
- Regan, V. (1995). Variation in French Interlanguage: A longitudinal study of sociolinguistic competence. In R. Bayley and D. R. Preston (eds.), *Second Language Acquisition and linguistic variation*, (pp. 177-201). Amsterdam: Benjamins.
- Rehner, K.; Mougeon, R. and Nadasi, T. (2003) The Learning of Sociolinguistic Variation by Advanced FSL Speakers: *The Case of Nous Versus On in Immersion French*. *Studies in Second Language Acquisition*, 25:127-156.
- Rickford, J. (1999). *African American Vernacular English*. London: Blackwell.
- Rickford, J. (2005). Against consensus: Challenging the contentions of the New Anglicists and divergencists concerning the development of AAVE. Paper

presented at New Ways of Analyzing Variation conference [NWAV 34], New York University.

- Rissanen, M. (1991). On the History of *That/Zero* as Object Clause Links in English. In Aijmer, K. and B. Altenberg (eds.), *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. (pp. 272-289). London: Longman.
- Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M. & Gagnon, L. (1997). Variation and the use of discourse markers in a language contact situation. *Language Variation and Change* 9(2):191-218.
- Schneider, E.W and Montgomery, M. (2001). On the trail of early nonstandard grammar: An electronic corpus of Southern U.S. antebellum overseers letters. *American Speech* 76(4):388-410.
- Seidlhofer, B. (2001). Closing a conceptual gap: the case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, 11(2):133-158.
- Seidlhofer, B. (2005). English as a lingua franca. In Wehmeier, S. (ed.) *Oxford Advanced Learner's Dictionary*. (p. 92). Oxford University Press
- Seidlhofer, B. (2006). *Vienna-Oxford International Corpus of English, Research Perspectives*. Retrieved December 1, 2006 from:  
[http://www.univie.ac.at/voice/voice.php?page=research\\_perspectives](http://www.univie.ac.at/voice/voice.php?page=research_perspectives)
- Smith, J. (2000). *Synchrony and diachrony in the evolution of English: Evidence from Scotland*. PhD, University of York
- Soars, L. and Soars, J. (1987). *Headway upper-intermediate*. Oxford: Oxford University Press.
- Spencer, T. (1999). *The Golden Grammar Book* (second edition). Lausanne, Switzerland: Imprivite S.A.
- Svartik, J. and Quirk, R. (1980). *A Corpus of English Conversation*. Lund: Gleerup.
- Tarone, E. (1988). *Variation in Interlanguage*. London: Edward Arnold.
- Tarone, E. (2002). Frequency effects, noticing, and creativity: Factors in a variationist interlanguage framework. *Studies in Second Language Acquisition* 24:287-296.
- Tagliamonte, S. (1997). Grammaticalization in apparent time: Tracing the Pathways of “gonna” in the city of York. Paper presented at Québec City, Canada.

- Tagliamonte S. (2002). Comparative sociolinguistics. In J. Chambers, P. Trudgill & N. Schilling-Estes (eds), *Handbook of Language Variation and Change* (pp.764-787). Oxford: Blackwell.
- Tagliamonte, S. and Smith, J. (2005). No momentary fancy! The zero complementizer in English dialects. *English Language and Linguistics* 9(2):289-309.
- Tagliamonte, S., Smith, J. and Lawrence, H. (2005). No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change* 17(1): 75-112.
- Thompson, S. and Mulac, A. (1991). The discourse conditions for the use of complementizer that in conversational English. *Journal of pragmatics* 15:237- 51.
- Todd, L. (1991). *Pidgins and creoles*. London: Routledge and Kegan Paul.
- Torres-Cacoullos, R. and Walker, J. (2003). Taking a complement ... variably. Paper presented at NWAVE 32, University of Pennsylvania.
- Tottie, G. and Harvie, D. (2000). It's All Relative: Relativization Strategies in Early African American English. In: S. (ed.). *The English History of African American English*. Oxford: Blackwell, 198-230.
- Tottie, G. and Rey, M. (1997). Relativization strategies in Earlier African American Vernacular English. *Language Variation and Change* 9: 219-247.
- Trudgill, P. (2004). *New Dialect Formation: The Inevitability of Colonial Englishes*. Edinburgh University Press.
- Trudgill P. and Hannah J. (1982). *International Englishes*. London: Arnold.
- Van den Eynden, N. (1993). *Syntactic variation and unconscious linguistic change: A study of adjectival relative clauses in the dialect of Dorset*. Frankfurt am Main: Peter Lang.
- Warner, A. (1982). *Complementation in Middle English and the Methodology of Historical Syntax*. London and Canberra: Croom Helm.
- Watts, R. J. & Murray, H. (eds.) (2001). *Die fünfte Landessprache? Englisch in der Schweiz*. Zürich: Hochschulverlag.
- Weinreich, U., Labov, W. & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics: A symposium*. Austin: University of Texas Press. pp. 95-188.

- Whinnom, K., (1971). Linguistic hybridization and the 'special case' of pidgins and creoles, in Dell Hymes (ed.), *Pidginization and Creolization of Languages, Proceedings of a conference held at the University of the West Indies Mona, Jamaica, April 1968*, (pp. 91-115). Cambridge University Press.
- Young, R. and Bayley, R. (1996). VARBRUL Analysis for Second Language Acquisition Research. In R. Bayley, and D. R. Preston (eds.), *Second Language Acquisition and Linguistic Variation* (pp. 253-306). Amsterdam: Benjamins.